MUSI 690

M.A. Thesis

An itemised study of the relationship between student attitude and performance as measured by selected musical tasks in the 2000 National Education Monitoring Project (NEMP) scores and survey results.

Name: Vanessa Tsang

Student Number: 53456694

Paper submitted in fulfilment for the degree of

Master of Arts,

School of Music

University of Canterbury

(2010)

Acknowledgements

The writer is grateful to her supervisor, Dr. Roger Buckton and Associate Professor David Sell, for their invaluable assistance in the preparation of this thesis. In particular, their patience and support will be fondly remembered and deeply appreciated.

Table of Contents

List of tables

List of figures

Abstract

The purpose of this study is to test the hypothesis that there is a link between children's musical experiences and attitudes, and their musical skills and understandings. For the purpose of the research, experience and attitudes have been measured by the National Education Monitoring Projects (NEMP) 2000 Music Survey, whilst the NEMP 2000 Assessment Results have been used to provide levels of musical skills and understandings. The first part of this study focuses on the nature of NEMP music assessment in relation to different aptitude, ability and achievement tests, and compares NEMP with various national assessments, in particular the ones used by the United States of America and Wales.

The second part of the study tests the hypothesis that whether a link is present between children's musical experiences and attitudes, and the extent of their musical skills and understandings, as revealed in their NEMP survey results and scores. The analysed results should be helpful in teacher education situations in determining the balance and relationship between teaching musical skills and understandings alongside student attitudes and previous experiences.

Based on the sample used in this research, correlation analyses showed that attitudes toward music do not have a strong influence on students' musical abilities. A greater number of relationships was found between musical experiences and students' NEMP scores. This result is similar to a few other attitude researches which will be briefly discussed and compared.

Chapter I - Introduction

1.1 Rationale

Improving the quality of education is not only in the minds of educationists and researchers, but in most cases, parents, and even students themselves. Throughout the years, different aspects of education have been evaluated in an attempt to provide an answer for better education. With the intention of measuring progress, as evidence of improved methods, some believed it is necessary to distinguish between innate talent and acquired abilities, while others argue otherwise. This argument over the concept of music ability has persisted through the course of the twentieth century, and has become divided into two major approaches – namely, the Gestalt and the Atomist approaches. Countries have based their national assessments from these approaches and in some cases created a range of educational testing methods and measurements.

The National Education Monitoring Project (NEMP) of New Zealand is one of the more recently developed national assessments in practice. The Educational Assessment Research Unit (EARU), financed by the Ministry of Education in New Zealand has developed NEMP as an integral part of a comprehensive national assessment strategy. This development attempts to assist the understanding of achievement standards and the quality of education in New Zealand in a wide range of curriculum areas. NEMP in music examines achievements related to creating music, re-creating music, responding to music and understanding music. It aims to measure student's ability, taking into consideration that achievement and talent complement each another. It attempts to demonstrate what New Zealand students can achieve and potentially assists in identifying issues such as the effectiveness of the curriculum or teaching methods. NEMP also has one unique element, which many other systems of

national assessment have not yet developed - the measuring of attitude.

Most educators would agree that, for children to learn effectively, interest plays a major part. William J. Bullock writes that the "frequent failure of music appreciation courses are due to insufficient recognition of the role of student['s] attitudes in acquiring appreciative ability."[1] Research by Broquist revealed that the favourableness of students' attitudes towards music steadily decreases as students' progress through the various school grades.[2] While Bullock stresses the importance of student's attitude in acquiring appreciation ability, Broquist suggests that the favourableness of students' attitudes is diminishing through the school years. Is the education sector giving sufficient recognition to the importance of student attitude might be the immediate question, however, more fundamentally, the question that needs to be addressed is, what is the relationship between attitude and ability?

> Since learning is so largely a matter of motivation, and the desire to learn, perhaps the first tests should be concerned with these factors, with attitudes and values, and other personality traits involved in the will to learn.[3]

Other studies have questioned the link between attitude and achievement. Tyler explained the extent to which correct identification of composer style is related to the degree of expressed pleasure obtained from music. He concluded that there is a "complete lack of relationship between liking … music and accuracy of

---

[1] Bullock, W. J. (1975). Student Attitude and Musical Response, *Journal of Aesthetic Education, 9*(4), 110.

[2] Broquist, O. A. (1961). *A Survey of the Attitude of 2,594 Wisconsin Elementary School Pupils towards their Learning Experience in Music,* Doctoral dissertation, University of Wisconsin, Madison, WI.

[3] Mueller, K. H. (1956). Studies in Music Appreciation, *Journal of Research in Music Education, 4*(1), 3-25.

identification."[4] A study of Michigan University students completing a course in 'effective living exhibited' found students change "in amounts of prejudice and in degree of personal adjustment … these changes were unrelated to growth in knowledge of the course."[5]

This raises the question as to whether there is a link between interest in music and the effectiveness of learning and appreciating music, is it purely that a child's aptitude in music reflects their learning? If there were a link between the two, it would be logical to create a curriculum that emphasises not only content, but also one which places greater emphasis on using effective methods to foster children's interests in music.

In the next chapters, there will be discussions on whether there is an existing relationship between attitude and music ability, and the reason for including attitude in the NEMP test of New Zealand; its uniqueness and a comparison of it with other countries' national assessments; the debate over the measurement of music ability, and problems involved in measuring attitude. As noted, this research is based on the scores and survey results from NEMP's year 2000 assessment. The background and function of NEMP is the focus of the next section.

---

[4] Tyler, L. E. (1946). An Exploratory Study of Discrimination of composer Style, *The Journal of General Psychology,* 34, 163.

[5] Mayhew, L. B. (1958). And in Attitudes. In Dressel, P. L. (Ed.) *Evaluation in the Basic College at Michigan State University,* New York: Harper. 226.

1.2 NEMP – the function of NEMP

National Education Monitoring Project (NEMP) is a vital part of a comprehensive national assessment strategy, which the Government initiated in an attempt to provide more information on achievement standards of New Zealand children, which are influence by in and out of schools factors, and the quality of New Zealand education in general. NEMP is funded by the Ministry of Education to monitor student progress over time in all curriculum areas. It aims to demonstrate students' abilities, and knowledge, and attempts to show how they apply what they have learned inside and outside schools, by completing tasks involving different levels and areas of difficulties. It was developed with the belief that it has the potential to contribute successfully to the improvement of learning in schools. NEMP results are considered by a national forum of curriculum and assessment specialists, principals, teachers, advisers and representatives of national educational organisations. The music report provides a useful insight into the present state of music education in our primary schools.

> The main goal of national monitoring is to provide detailed information about what children can do so that patterns of performance can be recognised, successes celebrated, and desirable changes to educational practices and resources identified and implemented.[6]

"Various government appointed commissions and working parties have drawn attention to the need for a systematic national information on educational results,"[7] because of this, NEMP was developed to use in primary education and implemented in New Zealand to try to answer these questions.

---

[6] Ministry of Education, EARU. (1997). NEMP Music Assessment Results 1996. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L. 5.

[7] Ibid.

The New Zealand Curriculum Framework states that the purpose of national monitoring is "to provide information on how well overall national standards are being maintained, and where improvements might be needed."[8]

In the NEMP reports, the purpose of national monitoring is stated as following:

1. To meet public accountability and information requirements by identifying and reporting patterns and trends in educational performance.
2. To provide high quality, detailed information which policy makers, curriculum planners and educator can sue to debate and review educational practices and resourcing.[9]

NEMP functions and aims are explained further by the Educational Assessment Research Unit (EARU) published in *NEMP Forum Comment*, which states that the focus is on the educational achievements and attitudes of New Zealand primary and intermediate school children.

<u>Design of NEMP and brief discussion on different behavioural objectives[10] of tests</u>

NEMP provides,

> a national 'snapshot' of children's knowledge, skills and motivation, and a way to identify which aspects are improving, staying constant, or declining. This

---

[8] New Zealand Ministry of Education. (1993). *New Zealand Curriculum Framework,* Wellington, N.Z.: Author. 26.

[9] Ministry of Education, EARU. (1997). NEMP Music Assessment Results 1996. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L. 5.

[10] Popham, W. J. (1978) *Criterion-referenced Meausrement.* Englewood Cliffs, NJ: Prentice-Hall. 14.

Behavioural objective is "a description of a performance you want learners to be able to exhibit before you consider them competent. An objective describes an intended result of instruction, rather than the process of instruction itself." Mager, R. (1975). *Preparing Instructional Objectives* (2nd ed.) Belmont, CA: Fearon-Pitman Publishers, Inc.

> information allows successes to be celebrated and priorities for curriculum change and teacher development to be debated more effectively, with the goal of helping to improve the education which children receive.[11]

From the goal embarked by NEMP, it is clear it would not be appropriate, to release information about individual students or schools. The aim is to encourage success rather than punish the underachievers. The crucial sole purpose is to provide information on which aspects of students' knowledge, skills and motivation are improving, staying constant, or declining. Because of these reasons, NEMP is considered as a low-stakes, criterion-referenced test.

Low-stakes testing has no consequences outside the school. Formative assessment is a good example of low-stakes testing. On the other hand, the results of high-stakes tests have important implications for the individual test taker.[12] Many criterion-referenced tests are also high-stakes tests, but, being a high-stakes test is not specifically a feature of a criterion-referenced test. It is instead an attribute of how an educational or government agency chooses to use the results of the test. American College Testing Programme (ACT) examination in USA is an example of criterion-referenced test without a cutscore that simply reflects students' knowledge at high school levels. Still, many criterion-referenced tests use a cutscore to determine whether the examinee has 'mastered' a specified level of the subject. NEMP demonstrates individual student's ability to perform selected tasks representative of ability criterions, thus criterion-referenced.

---

[11] Ministry of Education, EARU. 2005). NEMP Music Assessment Results 2004. (*National Education Monitoring Report 32.).* Dunedin, NZ.: Crooks, T. and Flockton. L. 5.

[12] Resnick, B. (2004). Majority of districts/schools employ "high-stakes" testing. *Successful School Marketer*. Retrieved December 9, 2005, from http://www.schooldata.com/ssm-resnick-majority.htm

Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement, Issues and Practice*, *20*(4), 19-28.

Before criterion-referenced tests, the more regularly used test method was the norm-referenced tests. The *Iowa Tests of Basic Skills*, more commonly known as the Iowa Tests is a well-known example of a norm-referenced test. They are standardised tests administered by the University of Iowa and used in the United States. K-12 students are tested on various subjects and are compared with his/her percentile score to overall national results. The Iowa Tests were first administered in 1935 as the *Iowa Every Pupil Tests*. The tests are mostly interpreted norm-referenced, to provide grading. Still, the Iowa Tests can be interpreted criterion-referenced, as long as the user has established some performance standards against which comparisons to be made. The primary difference between criterion-referenced tests and norm-referenced tests lies in how the results are used.

Scotland and France both have low-stakes testing like NEMP. The assessment results are used within schools for monitoring progress and standards and by the inspectorate for external review, but these uses, according to Crooks, "do not have a dominant influence on teachers' assessment practice."[13] On the other hand, he stated that assessment in New Zealand primary schools "is predominantly low-stakes assessment focused on monitoring pupils' learning, improving learning through direct feedback to students or adjustments to teaching programmes. Written or oral reports to parents can be seen as complementing the formative role by giving guidance to parents and students, while also having a summative role."[14]

---

[13] Crooks, T. J. (2002). Educational Assessment in New Zealand Schools, *Assessment in Education, 9*(2), 246.

[14] Ibid.

New Zealand is not alone in the step of moving teachers away from dependence on test-derived data. The change in Wales' *Curriculum 2000*[15] also moves teachers away from test-derived data with emphasise on gaining more understanding and having a higher interaction level with their students. NEMP's importance on an international level will be discussed further in chapter 2.3.

The detailed discussion on the use of different testing method is beyond the scope of this study, but a brief definition of the different methods of test result measurement can be found at the end of this chapter.

After establishing the focus of NEMP being a low-stake, criterion-referenced test, we now look at the possible origin of the idea of NEMP.

Idea behind NEMP

In order to consider achievement standards and the quality of education of a country, there are differing views on methods of assessment. In *Assessing Achievement Versus High-Stakes Testing: a Crucial Contrast*, Lyle V. Jones discusses Tyler's view of assessment. 'Assessment'[16] was the term proposed by Tyler to represent the achievements evaluation of groups of people with different ages, perhaps also people with different demographic, ethnic background, or geographical differences.

Tyler distinguished assessment from three other forms of educational appraisal:

---

[15] ACCAC (Qualifications, Curriculum and Assessment Authority for Wales) (2004). *Review of the School Curriculum and Assessment Arrangements 5-16.* Cardiff: ACCAC.

[16] Tyler, R. W. (1966). The objectives and plans for a National Assessment of Educational Progress. *Journal of Educational Measurement, 3*(1), 1-4.

(a) Testing individuals to assign grades or to select students for further opportunities (with high-stakes consequences for the individuals),

(b) Diagnosing students to plan subsequent teaching, and

(c) Evaluating the effectiveness of a curriculum or a set of teaching methods.[17]

Jones said the purpose of assessment is to "give dependable information about population and subpopulation's progress in levels of educational achievement over time."[18] This is not very different from NEMP's goal to provide information on New Zealand primary education. The fact that the results of NEMP are called assessment results might also suggest the possible relevance.

Before NEMP, the National Assessment of Educational Progress (NAEP), the national assessment system used in the United States, was first to be in accordance with Tyler's idea of assessment. NAEP was designed to assess a national sample of students.[19] Students would take no more than a fraction of the exercises, and no score would be obtained from any student's performance. Many of the exercises, in the form of hands-on problems would be solved, while some exercises involve a consensus derived by a group of children, would represent a broad range of difficulty and a full range of educational objectives in ten different subject areas. Professional administrators were trained to provide highly controlled assessment conditions. Exercises were read aloud, so that deficiencies in reading would not prevent good performance in different areas, for example, mathematics. An 'I don't know' alternative was offered to discourage guessing and to reduce non-response. Detailed

---

[17] Ibid.

[18] Jones. L. V. (2001). Assessing achievement versus high-stakes testing: a crucial contrast. *Educational Assessment*, *7*(1), 1.

[19] Ibid. 23.

results of periodic assessments for four different age groups (9, 13, 17 years and young adults) were reported. Jones believed this way the public would have concrete evidence about what respondents knew and could do, and about changes in performance over time. Up to this point, the similarities between the NAEP and NEMP are already evident.

NAEP became operational in 1969, 24 years ahead of NEMP. Since 1969, NAEP has not only survived but has been expanded. Still, Jones noted many of the original objectives were changed, if not, compromised. First, the more traditional multiple-choice and short-answer questions have replaced the original, desired, rich variety of exercises. Exercises became consistent in difficulty. The young adult sample was eliminated, and school grade replaced age as the primary unit of assessment. The ten subject areas have received uneven attention with mathematics, reading, science, and writing assessed far more often than literature, social studies, art, music, citizenship, and career development. Exercises are no longer read aloud, and the "I don't know" alternative has been removed. Because of the state-by-state assessments that has taken its place in 1990, professionally trained administrator were replaced by local school personnel for convenience. This has raised questions about the uniformity and consistency of administration.[20]

Perhaps, one of the major changes in NAEP would be the measurement method. It has shifted from an assessment that is more criterion-referenced to a more

---

[20] Jones. L. V. (2001) Assessing achievement versus high-stakes testing: a crucial contrast. *Educational Assessment*, *7*(1), 23.

norm-referenced assessment. Instead of reporting a percentage-correct score for each exercise, scale scores were developed to simplify the procedure of reporting large clusters of exercises. More recently, the scale scores have developed into achievement levels to compare actual performance with how good performance 'should be'. "Although providing a more popular basis for interpreting results, arbitrary procedures for establishing basic, proficient, and advanced performance have led to controversy."[21] The fact that an element of how good performance 'should be' is added would suggest a more norm-referenced measurement with a high-stake implication. When parents, teachers or students are lead to believe how good performance 'should be', the assessment becomes subjective with possible negative psychological effect for the 'less talented'.

Item response technology is used. Although different children take different sets of exercises, scores are inputted for each child in the sample. Inputted scores then are averaged for any specified subgroup of children. Jones believes, although many of these changes were well intentioned, with some clearly supported by psychometric considerations, and the need to achieve better public communications, nonetheless, many of the changes are no longer congruent with Tyler's (1966) concept of assessment.[22]

---

[21] National Academy of Education. (1993). Setting performance standards for student achievement. *U.S. General Accounting Office*. Stanford, CA: Author.

[22] Jones. L. V. (2001) Assessing achievement versus high-stakes testing: a crucial contrast. *Educational Assessment*, *7*(1), 24.

After 24 years, Jones said it is noteworthy that Tyler's views on assessment seems to have materialised in New Zealand's National Education Monitoring Project (NEMP). For the primary purpose of NEMP is "to meet public accountability and information requirements by identifying and reporting trends in educational performance ... [and] ... to provide high quality, detailed information which policy makers, curriculum planners and educators can use to debate and review educational practices and resourcing."[23] The 2004 NEMP assessment report stated:

> National Education Monitoring Project commenced in 1993, with the task of assessing and reporting on the achievement of New Zealand primary school children in all areas of the school curriculum. Children are assessed at two class levels: Year 4 (halfway through primary education) and Year 8 (at the end of primary education). Different curriculum areas and skills are assessed each year, over a four-year cycle.[24]

The following section discusses the implementation of NEMP assessments.

Implementation of NEMP

Each year, a small random sample of Year 4 (age 8-9) and Year 8 (age 12-13) are selected to represent about 3 percent of the children at those levels in New Zealand schools. (In 2000 and 2004, 2.5 percent of the children at school were selected.) Subsequently, these children are assessed in their own schools by teachers specially chosen and trained for this work. The selected teachers usually work in their own region of New Zealand. These teachers are selected from a national pool of applicants, who have attended a week of specialist training in Wellington led by senior project

---

[23] Ministry of Education, EARU. (1997) NEMP Music Assessment Results 1996. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L. p5.

[24] Ministry of Education, EARU. (2005) NEMP Music Assessment Results 2004. (*National Education Monitoring Report 32.).* Dunedin, NZ.: Crooks, T. and Flockton. L. p2.

staff. Then, they work in pairs to conduct assessments of 60 children over five weeks. Their employing school is fully funded to employ a relief teacher during their secondment.

Task instructions are given orally by teachers following written text, through video presentations, or in writing. Many of the assessment tasks involve the children in the use of equipment and supplies. Their responses are presented orally, by demonstration, in writing, or through submission of other physical products. Many of the responses are recorded on videotape for subsequent analysis.

In 1996, three task approaches were used for the music assessment. Each student was expected to spend about an hour working in the first two formats and two one-hour sessions working in the third format.

Different students attempt different tasks. The 1440 students selected at each level are divided into three groups of 480 students, comprising four students from each of 120 schools.

The assessments take place in the second half of the school year, between August and November, The year 8 assessments occur first, over a five week period. The year 4 assessments follow over a similar period. Each student participates in about four hours of assessment activities spread over one week.

Four assessment approaches are in use, but only three approaches were used in 1996 for music assessment:

One-to-one interview: Each student worked individually with a teacher with the whole session recorded on videotape.

Station: Four students, working independently, moved around a series of stations where tasks had been set up. This session was not videotaped.

Team: Four students worked collaboratively, supervised by a teacher, with the whole session recorded on videotape.[25]

In 2000 and 2004, a fourth approach was adopted in music, which had previously been used in other subjects:

Group and Independent: Four students worked collaboratively, supervised by a teacher, on some tasks. This was recorded on videotape. The students then worked individually on some paper-and-pencil tasks.[26]

A four-year assessment cycle is used for individual subjects. One third of the tasks were kept constant from one cycle to the next. This re-use of tasks allows trends in achievement across a four-year interval to be observed and reported. Starting from 2002, the percentage of tasks retained was increased from 35 to 45 percent, so that trends will be able to be reported more thoroughly.

Marking and administering

The marking and analysis of students' work occurs in Dunedin. Most of the tasks which can be marked objectively or with modest amounts of professional experience are marked by senior tertiary students, most of whom have completed three or more years of preservice preparation for primary school teaching. The student markers for the 1996 tasks were employed five hours per day for the period ranging between 5

---

[25] Ministry of Education, EARU. (1997) NEMP Music Assessment Results 1996. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L. 7

[26] Ministry of Education, EARU. (2005) NEMP Music Assessment Results 2004. (*National Education Monitoring Report 32.).* Dunedin, NZ.: Crooks, T. and Flockton. L. 7.

weeks and 9 weeks.

Incidentally, the marking and analysis of NEMP results have unintentionally acquired more benefits. Following previous section's discussion on the idea behind NEMP, Jones suggested NEMP aims to show national standards rather than focus on individual child or school, appears to be designed in keeping with Tyler's idea[27] of how assessments should be conducted. Assessment results receive a good deal of public attention, and the assessment design has constructively influenced school curricula and methods of teaching. As one side benefit, the project employs about 100 teachers a year to administer the assessments in the schools.[28] "Strong evidence has been assembled concerning the value of their experience to these participant teachers."[29]

The tasks which required higher levels of professional judgement are marked by teachers, selected from throughout New Zealand. In 1996 approximately two thirds of the teachers who applied were appointed, which was a total of 150. Most teachers worked either mornings or afternoons for one week. Their ratings of the experience were overwhelmingly positive, with 87 percent stating emphatically that the experience was "professionally satisfying and interesting". The teacher administrators reported that they found their training and assessment work very stimulating and professionally enriching. Several teachers have reported major changes in their

---

[27] Tyler, R. W. (1966). The objectives and plans for a National Assessment of Educational Progress. *Journal of Educational Measurement,* 3(1), 1-4.

[28] Jones. L. V. (2001) Assessing achievement versus high-stakes testing: a crucial contrast. *Educational Assessment*, *7*(1), 24.

29 According to Jones, L. V. (2001), Gilmore, A. (1999). NEMP: Evaluation of teachers' professional development. In M. Kane (Chair), The National Monitoring Project: Validity, professional development, and group assessment processes. Symposium conducted at the annual meeting of the American Educational Research Association, Montreal, Canada.

teaching and assessment practices because of their experience. Given that, 95 teachers served as teacher administrators in 1996, or about half a percent of all primary teachers, the project is making a major contribution to the professional development of teachers in assessment knowledge and skills. This contribution will steadily grow, since preference for appointment each year is given to teachers who have not previously served as teacher administrators.

Demographics

Although the emphasis is on the overall national picture, NEMP also pays attention to possible differences in performance patterns for different demographic groups and categories of school. The variables considered are:

- Student gender: male, female.
- Student ethnicity: Maori, non-Maori.
- Geographical zone: Greater Auckland, other North Island, South Island.
- Size of community: urban area over 100,000, community of 10,000 to 100,000, rural area or town of less than 10,000.
- Socio-economic index for the school: bottom three deciles, middle four deciles, highest three deciles.
- Percent of Maori children in the school: less than 10 percent, 10 to 30 percent, more than 30 percent.
- Percent of Pacific Island children in the school: less than 5 percent, 5 percent or more.
- Size of school:

Type of school (for year 8 sample only): Full primary school, intermediate school (some students were in other types of schools, but too few to allow separate analysis).

Table One

Size of Schools

Year 4 schools	less than 20 year 4 students	20-35 year 4 students	More than 35 year 4 students
Year 8 schools	less than 35 year 8 students	35-150 year 8 students	More than 150 year 8 students

The consideration of different demographics not only provides more variety of detailed information, it also increases the accuracy of the result interpretation.

Perception of NEMP by academics

NEMP has received many positive comments from scholars of different countries. In June 1996, Scholars from the United States and England, with distinguished international reputations in the field of educational assessment, accepted and invitation from the project directors to visit Dunedin. They conducted a thorough review of the progress of NEMP, with particular attention to the procedures and tasks used in 1995 and the results emerging. At the end of their review, they prepared a report which concluded as follows:

> The National Education Monitoring Project is well conceived and admirably implemented. Decisions about design, task development, scoring, and reporting have been made thoughtfully. The work is of exceptionally high quality and displays considerable originality. We believe that the project has considerable potential for advancing the understanding of and public debates about the educational achievement of New Zealand students. It may also serve as a model for national and/or state monitoring in other countries.[30]

---

[30] Professors Paul Black, Michael Kane & Robert Linn. (1996), as cited in Ministry of Education, EARU. (1997) NEMP Music Assessment Results 1996. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L. 8.

A further review was conducted late in 1998 by another distinguished panel. Amid suggestions for further refinements and investigations, they commented that:

> We want to acknowledge publicly that the overall design of NEMP is very well thought through……The vast majority of tasks are well designed, engaging to students and consistent with good assessment principles in making clear to students what is expected of them.[31]

So far, this section has provided NEMP information that applies across the different assessed subjects, in the next section, the focus will be on music.

---

[31] Professors Elliot Eisner, Caroline Gipps and Wynne Harlen. (1998), as cited in Ministry of Education, EARU. (2001) NEMP Music Assessment Results 2000. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L. 9.

NEMP – Music aspects

The New Zealand Curriculum Framework states that the purpose of national monitoring is to "provide information on how well overall national standards are being maintained, and where improvements might be needed."[32]

NEMP aims to measure the achievement of New Zealand primary school children, thus, providing a national 'snapshot' of children's knowledge, skills and motivation, and whether New Zealand education has a balanced curriculum or needs further improvement.[33] A music report is written by Educational Assessment Research Unit (EARU) after each assessment to provide insights into the present state of music education in primary schools.

The NEMP assessment is unique to New Zealand, combining individual and group tests, with a four-year testing period and special link tests to compare the achievement of New Zealand children in different years; thus, the 1996, 2000 and 2004 NEMP assessment results will be the most important source of this research. Because of the unique individually designed nature of the NEMP test, taking into consideration of New Zealand culture, this research will focus more on the NEMP test rather than the other tests.

Since the inception of NEMP assessment in 1996, the New Zealand curriculum has inevitably changed. The overall framework has a central organising theme supported

---

[32] New Zealand Ministry of Education. (1993). *New Zealand Curriculum Framework,* Wellington, N.Z.: Author. 26.

[33] Ministry of Education, EARU. (1997) NEMP Music Assessment Results 1996. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L. 5.

by three interrelated aspects: content, process, and motivation and involvement aspects. From 2000 to 2004, the music content aspect changed from creating, re-creating, appreciating music (based on the 1989 music syllabus)[34] to creating, performing, understanding and responding to music (based on the new *Arts Curriculum*)[35].

Currently, there are four published music NEMP assessment results, namely, 1996, 2000, 2004 and 2008 was released in June in 2009. The use of tasks with both year 4 and year 8 students allows comparisons of the performance of year 4 and 8 students in 2000. Because some tasks have been used twice, in 1996 and again in 2000, trends in performance across the four year period can also be analysed. The reason for this research to be based on the 2000 results rather than the most recent 2004 and 2008 results is that the link tasks, which are used to allow comparisons between the 2004 and 2008 assessments, were not yet published when this research started.

Experimental design factors and NEMP music results

The NEMP assessment results and surveys are used in this research to discover whether there is a relationship between children's musical attitudes and experiences, and the effectiveness of their musical performances and abilities.

The assessment tasks covered a wide range of musical activities including the following skills:

1. Listening,

---

[34] Department of Education. (1989). *New Zealand Syllabus for Schools of Music Education – Early Childhood to Form Seven.* Wellington, N.Z.: Author.

[35] Ministry of Education. (2000). *The Arts in the New Zealand Curriculum.* Wellington, N.Z.: Author.

2. Creating,
3. Moving,
4. Directing,
5. Singing,
6. Playing,
7. Reading and
8. Recording[36]

Although part of the NEMP test was done in groups and part individually, only the individual tasks are analysed in this research. Many researchers hold differing views on the value of one-to-one - individually assessed tasks (Students work individually with a teacher). Sue Braatvedt suggests, "for the purposes of the NEMP singing tasks, individual evaluations seemed to be most appropriate."[37] "This view is also held by Shuter-Dyson, that individual tests are usually more reliable because of the instructor's ability to clarify any instructions which might appear confusing or vague to the child."[38] Still, Braatvedt mentions that the disadvantage of individual tests is the time consuming factor and thus only a small sample of a population is obtained. Also, another disadvantage being that, it is very likely the child might feel vulnerable as the sole candidate of a one-to-one individual task. Consequently, the child may not respond well, and the results will be affected accordingly. However, Roger Buckton (1983)[39] designed a singing test he devised in which the whole class was able to participate. With the use of individual microphones, it made it possible to record individual's voices.

---

[36] Ministry of Education, EARU. (2001) NEMP Music Assessment Results 2000. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L. 10.

[37] Braatvedt, S. P. (1998). *The Role of Singing in the NEMP Tests.* Unpublished master's thesis dissertation, University of Canterbury, Christchurch, New Zealand. 17.

[38] Cited by Braatvedt, S. P. (1998). 16. Shuter-Dyson, R. (1968). *The Psychology of Musical Ability.* London: Methuen & Co Ltd. 20.

[39] Buckton, R. M. (1983). Sing a Song of Six-Year-Olds. *New Zealand Council for Educational Research,* Wellington. 10, 11.

The four tests analysed in this study focus mainly on keyboard and listening skills. They had the same problem as the singing tests. The video footage shows many children felt vulnerable as the sole candidate in the situation, which might have affected the results.

The advantage of individual tests were that they were bound to be more reliable because of the teacher administrator's ability to clarify any instructions, which might appear confusing, or vague to the child. Still, there was a significant variance in the administration of the tests by different teacher administrators. These factors will be discussed further in chapter 4.6, in the section on limitations of research.

1.3 Definition of terms

Ability
The quality or state of being able; power to perform, whether physical, moral, intellectual, conventional, or legal; capacity; skill or competence in doing; sufficiency of strength, skill, resources, etc.

Achievement
The act of achieving or performing; an obtaining by exertion; successful performance; accomplishment; as, the achievement of his/her object.

Aptitude
Natural ability to acquire knowledge or skill.

Attitude
Disposition or state of mind.

Talent
A marked ability or skill.

Musicality
The condition of being musical.

Understanding
Mental, sometimes emotional process of comprehension, assimilation of knowledge, which is subjective by its nature; reason or intelligence, ability to grasp the full meaning of knowledge, ability to infer.

Appreciating
To be fully conscious of; be aware of; detect.

Learning
Learning is divided into two general parts: discrimination and inference. D and I are not mutually exclusive.

Rote Learning
Discrimination learning is rote learning. It takes place when students are conscious of what they are learning, because they are being taught by a teacher. Inference learning

is conceptual learning. It takes place when students are not conscious of what they are learning, because they are teaching themselves what they are learning.

Evaluation
Evaluation is subjective, but it must be based upon objective measurement.

Rhythm
'Time in music', but is also used as a term for beat/pulse as in NEMP.

Rhythmic pattern
A rhythmic phrase on a single pitch which may be repeated as in an ostinato.

Melodic pattern
A rhythmic phrase which consists of changing pitch. It may be repeated as an ostinato.

Ostinato
A short musical pattern which is repeated over and over and may be at a single pitch (rhythmic ostinato) or several pitches (melodic ostinato).

Different testing methods:

High-stakes testing
A test which has important consequences for the test taker. If the examinee passes the test, then the examinee may receive significant benefits, such as a high school diploma, a scholarship, or a license to practice law. If the examinee fails the test, then the examinee may receive significant disadvantages, such as being forced to take remedial classes until the test can be passed, or not being allowed to drive a car. The key features of a high-stakes test are:

- A single defined assessment
- A clear line drawn between those who pass and those who fail
- A direct consequence for passing or failing (something 'at stake')

Low-stakes testing
Low-stakes testing has no consequences outside the school, although the results may have classroom consequences such as contributing to students' grades. Formative assessment is a good example of low-stakes testing.

Criterion-referenced tests

These tests are designed to show how students achieve in comparison to standards, usually state standards. (NASBE, 2001; Wilde, 2004; Zucker, 2003). In contrast to norm-referenced tests, it is theoretically possible for all students to achieve the highest—or the lowest—score, because there is no attempt to compare students to each other, only to the standards. Results are reported in levels that are typically basic, proficient, and advanced. The test items are not chosen to rank students but to ascertain whether they have mastered the knowledge and skills contained in the standards. Sometimes they are called standards-based tests—begins from standards, which list the knowledge and skills students are expected to learn. Because standards are usually far more numerous than could ever be included in a test, test designers work with teachers and content specialists to narrow down the standards to essential knowledge and skills at the grades to be tested. They are the basis for the development of test items.

Norm-referenced tests

These tests are designed to compare individual students' achievement to that of a "norm group," a representative sample of his or her peers. The design is governed by the normal or bell-shaped curve in the sense that all elements of the test are directed towards spreading out the results on the curve (Monetti, 2003; NASBE, 2001; Zucker, 2003; Popham, 1999). The curve-governed design of norm-referenced tests means that they do not compare the students' achievement to standards for what they should know and be able to do—they only compare students to other students who are assumed to be in the same norm group. Results of norm-referenced tests are frequently reported in terms of percentile. Percentile rankings are often used to identify students for various academic programs such as gifted and talented, regular, or remedial classes. On a symmetrical bell curve, a score in the 50th percentile is the average. Norm-referenced tests lead to frustration on two counts. First they frustrate the teacher's success in teaching important knowledge and skills because students are unlikely to face questions about that skill and knowledge on the test (Popham, 1999). Second, no group of students can achieve at higher levels without others achieving at lower levels. Norm-referenced tests make it mathematically impossible for "all the children to be above average" (ERS; Burley, 2002).

Standardised tests – Administration consistency

Standardised testing means that a test is "administered and scored in a predetermined, standard manner" (Popham, 1999). Students take the same test in the same conditions

at the same time, if possible, so results can be attributed to student performance and not to differences in the administration or form of the test (Wilde, 2004). For this reason, the results of standardised tests can be compared across schools, districts, or states. Standardised testing is sometimes used as a shorthand expression for machine scored multiple-choice tests. However, standardised tests can have almost any format.

Analysis of Variance (ANOVA)
A statistical technique that determines whether three or more means are statistically different from each other.

Chi-square analysis
Assesses how closely the observed frequencies fit the pattern of the expected frequencies and is referred to as a "goodness-of-fit" test.

Regression analysis
A statistical technique that analyses the linear relationship between two variables by estimating coefficients for an equation for a straight line. One variable is designated as a dependent variable and the other is called an independent or predictor variable.

Correlation
In probability theory and statistics, correlation indicates the strength and direction of a linear relationship between two random variables. That is, in contrast with the usage of the term in colloquial speech, which denotes any relationship, not necessarily linear. In general statistical usage, correlation refers to the departure of two random variables from independence. In this broad sense there are several coefficients, measuring the degree of correlation, adapted to the nature of the data.

Pearson correlation coefficient
A statistical measure of the strength of a linear relationship between two metric variables

Method and Technique
A method is the order in which sequential objectives are introduced in a course of study to accomplish a comprehensive objective. A technique is a teaching aid which is used to achieve a sequential objective. Method refers to 'why' we teach 'what' we teach 'when' we teach it. Technique refers to 'how' we teach it.

References for definition of terms:

Burley, H. (2002). A Measure of Knowledge. *American School Board Journal*, *18*(2).
Educational Research Service. (n.d.). *Focus on high-stakes testing*. Arlington VA: Educational Research Service

Gordon, E. E. (1980). *Learning sequences in Music: Skill, Content, and Patterns*. (1984 edi.). Chicago: IL: GIA.

Hair et al. (2003) *Marketing Research – Within a changing Information Environment* (2nd ed.). New York, NY: McGraw-Hill.

Mitchell, R. (2006, Feb. 15). Research review: Effects of high-stakes testing on instruction. In *The Center for Public Education.* Retrieved Nov. 27, 2008 from http://www.centerforpubliceducation.org/site/c.goJQI0OwElH/b.2119025/k.C92D/The_nature_of_assessment_A_guide_to_standardized_testing.htm

Monetti, D. M., & Hinkle, K. T. (2003). *Five important test interpretation skills for school counselors*. ERIC Digest. ED481472 2003-09-00.

National Association of State Boards of Education. (2001). *A primer on state accountability and large-scale assessments*. Available: http://www.nasbe.org/Educational_Issues/Reports/Assessment.pdf

Popham, J. W. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, *56*(6), 8-15.

Webster, N. (1913). *Webster's Revised Unabridged Dictionary.* Cambridge, MA: G & C. Merriam Co.

Wilde, J. (2004). Definitions for the no child left behind act of 2001: Assessment. Washington DC: National Clearinghouse for English Language Acquisition (NCELA).

Zucker, S. (2003). *Fundamentals of standardized testing*. San Antonio TX: Harcourt Assessment, Inc.

Chapter II – Review of related studies

The characteristics of NEMP assessment can be attributed to a range of previous research related to testing various populations of children. In this chapter, major designs of tests, the influence of attitude on learning will be reviewed in chapters 2.1 and 2.2. In chapter 2.3, the uniqueness of NEMP will be discussed in relation to other systems of national testing.

2.1 Music ability tests

Many influential studies, such as those by Seashore(1938)[40], Drake(1957)[41], Wing(1948)[42], Gordon(1987)[43], and Bentley(1966)[44], among others, were concerned with the design of formalised tests to provide scored ratings of comparative musicality. These are all norm-based tests. Norm-referenced and criterion-referenced testing are briefly discussed in chapter 1.2 in relation to NEMP, and their definitions can be found in 1.3.

The tests can be divided into two groups, separated by the two general points of view towards the description of music aptitude: Gestalt and Atomistic views.

---

[40] Seashore, C.E. (1938). *Psychology of Music.* New York: McGraw-Hill.

[41] Cited in Shuter-Dyson, R., & Gabriel, C. (1981). *The Psychology of Musical Ability* (2nd edi.). London: Methuen.

[42] Wing, H. (1948) *Tests of Musical Ability and Appreciation.* London: Cambridge University Press.

[43] Gordon, E. E. (1987). *The Nature, Description, Measurement, and Evaluation of Music Aptitudes*. Chicago: IL: GIA.

[44] Bentley, A. (1966). *Musical Ability in Children and its Measurement.* New York: October House.

Gordon(1987)[45] is the representative Gestalt view, and Bentley(1966)[46] the latter. Gordon, along with theorists Wing, Mursell, and Shuter believe that music ability is an all-inclusive or all-pervasive trait, and comprises more than a set of specific attributes dependent upon sensory capacities, while Seashore, Mainwaring and Bentley are representatives of the 'atomistic' approach. Lying between these extreme viewpoints are theories such as Drake and Holmstrom. Gordon's view on the two viewpoints were expressed in *Learning Sequences in Music*,

> The Gestalt group holds that music aptitude is a unitary trait of which general intelligence is a substantial part. The atomistic group argues that music aptitude is multidimensional, that it has various parts, and that none is significantly related to general intelligence. Just what all of the parts are remain unknown. Nevertheless, both schools of thought make provision for the measurement of tonal, rhythmic, and aesthetic – expressive – interpretive qualities, either separately or collectively, and it is agreed that music aptitude is manifested in preferences as well as objectivity.[47]

In Seashore's terminology, an individual's response to the acoustic features of sound, whether presented singly or in combination, was assessed by itemised psychometric tests utilising sound's psychological correlates. He claimed that musical capacity may be divided into varying degree of talents and based these talents on the properties of sound: pitch, volume, timbre and duration. The resultant data was believed to reveal some underlying musical 'capacity', 'capability', or 'talent', and to be a predictor of musical 'aptitude'. Seashore holds the atomist view that music aptitude evaluation is something we must "patiently settle down to the isolation and observation of isolable

---

[45] Gordon, E. E. (1987). *The Nature, Description, Measurement, and Evaluation of Music Aptitudes*. Chicago: IL: GIA.

[46] Bentley, A. (1966). *Musical Ability in Children and its Measurement.* New York: October House.

[47] Gordon, E. E. (1980). *Learning sequences in Music: Skill, Content, and Patterns*. (1984 edi.). Chicago: IL: GIA. 226.

traits."[48] Most investigators who have studied the inter-correlations of Seashore's *Measures of Musical Talents* tests have found low correlation between his six tests, namely, Pitch, Loudness, Rhythm, Time, Timbre, and Tonal Memory. This supports Seashore's view that the differing music talent traits are relatively independent.

Gestalt believers have criticised Seashore's view on the grounds that since music involves patterns and relationships, evaluation of music should not be the discrimination of the basic constituents of sound. Mursell argued that music does not rely on the stimuli reaching the outer ear and the response of the inner ear, but on the organising and transforming operation of the brain.[49]

Herbert Wing's *Standardised Tests of Musical Intelligence* (Wing, 1948)[50] is considered by many to be the antithesis of Seashore's approach. Wing believed the seven component tests of chord analysis, pitch discrimination, memory, rhythmic accent, harmony, intensity, and phrasing may be combined into an overall measure of music ability. However, the objectivity in his research was in question, Buckton observed that, for example, in the test of tonal memory, notes of unequal length may have been chosen instead of a more rhythmic rhythm to enhance the independence of

---

[48] Seashore, C. E. (1938). *Psychology of Music.* New York, McGraw-Hill. 332.

[49] Mursell, J.L. (1937). *The Psychology of Music.* New York, W.W. Norton and Co. 51.

[50] Wing, H. (1948). *Standardised Tests of Musical Intelligence.* National Foundation for Educational Research. Upton Park, Slough.

the seven tests. "It seems that test constructors own beliefs influence test design, which in turn strengthen beliefs!"[51]

Kirchhubel constructed a list of different tests grouped under different dimensions of music that have been frequently tested by different authors:

- audio-acoustical perception (Seashore, Kwalwasser and Dykema, Tilson, Drake, Bentley),
- tonal concepts (Seashore, Kwalwasser and Dykema, Wing, Tilson, Gaston, Kwalwasser, Drake, Gordon, Bentley),
- rhythm concepts (Seashore, Kwalwasser and Dykema, Gordon, Bentley), and
- expressive-interpretive (aesthetic) concepts (Seashore, Kwalwasser and Dykema, Wing, Gaston, Gordon).[52]

It is interesting to note that while Gordon and Bentley have both tested tonal and rhythm concepts of music, Bentley tests audio-acoustical perception while Gordon tests expressive-interpretive (aesthetic) concepts.

Gordon's *Musical Aptitude Profile* (Gordon, 1965)[53] is probably his earlier more well-known work. The musical aptitude profile is divided into three main sections – Tonal Imagery, Rhythm Imagery, and Musical Sensitivity. All tests use especially composed music, which involved making comparisons between two musical extracts.

---

[51] Buckton, R. M. (1981). *The Development of Musical Concepts in Young Children: an Investigation.* Unpublished doctoral dissertation. University of Auckland, Auckland, New Zealand. 24.

[52] Kirchhubel, J. (2002). *Adolescent Music Development and the Influence of Pre-Tertiary Specialised Music Training.* Unpublished doctoral dissertation. Griffith University, Brisbane, QSL, Australia. 31.

[53] Gordon, E. E. (1965). *Musical Aptitude Profile.* Boston: Houghton Mifflin.

In Gordon's *Music Aptitude* tests, the consistent use of the musical setting underlines his rationale that one intuitively derives musical understanding from organised sound, rather than how well one may give organised sound some theoretical explanations; observing musical aptitude in a musical setting instead of dissecting music into individual independent sound elements and thereafter observe accordingly. In his 1989 work, he described music aptitude as a measure of a student's 'inner possibility', and explained music achievement as a student's 'outer possibility', the measure of what a student has learned.[54]

Lyman has commented that Gordon's *Music Aptitude* tests' high levels of validity and reliability "may have been achieved partly at the expense of usability"[55], a total administration time of 150 minutes.

After Gordon's 1965 publication of *Musical Aptitude Profile* in U.S.A., Bentley's *Measures of Musical Abilities* were published in the U.K. the following year. The Bentley test is similar to Seashore and Wing aural acuity tests, with considerable shorter test duration to suit a younger age group, but with reliability at a more moderate level. It gained its popularity by the size of the test, with the whole battery test lasting only twenty minutes, and the relative ease of administration.

---

[54] Cited by Kirchhubel, J. (2002). *Adolescent Music Development and the Influence of Pre-Tertiary Specialised Music Training.* Unpublished doctoral dissertation. Griffith University, Brisbane, QSL, Australia. 26.

[55] Lyman, H. B. (1971). *Test Scores and What They Mean.* Englewood Cliffs, New Jersey: Prentice-Hall. 21.

The data of these studies is used by many music researchers to justify why some people achieve limited success in music; and by using this data as a basis, other music researchers have argued the lack of consideration on the effect of training versus innate talent.

Consequently, the problem lies not only within the decision of which test or tests to use, these researchers find we need to question several things: is the test really testing aptitude and not achievement? Is it true that musical aptitude becomes, at a relatively early age, noncumulative? Some tests were developed to try to focus on the measuring of student achievement rather than aptitude.

Music Achievement Tests by Richard Colwell (1970) are divided into four independent tests:

1. Pitch, interval and meter discrimination,
2. mode and auditory visual discrimination and tonal centre identification,
3. melody, pitch and instrument recognition and tonal memory,
4. musical style, auditory-visual discrimination, and chord and cadence recognition.[56]

He believes the tests enable teachers to determine how well each student has mastered the basic auditory object of the school music programme. By measuring whether a student has mastered a certain task, the test avoids considering the musical aptitude of the student, but demonstrates what the student is capable of achieving. In other words, showing the student's natural aptitude as well as training given. Colwell regards scores in his achievement tests as dependent upon a student's musical aptitude plus qualities of perseverance, interest, and self discipline. This latter point causes Colwell

---

[56] Colwell, R. (1970). *Interpretive Manual Music Achievement Tests 1 and 2*. Chicago: Follett. 26.

to conclude that since these qualities are also required for success in music, "an achievement test such as MAT would logically have more predictive value than administration of an aptitude test alone."[57]

The purpose of Colwell designing an achievement-oriented test was to be able to consider both learning and innate abilities, rather than aptitude alone. This is similar to what the NEMP results used in this research has tried to achieve. It was discussed in chapter one that NEMP's aim is "to provide detailed information about what children can do, so that patterns of performance can be recognised, successes celebrated, and desirable changes to educational practices and resources identified and implemented."[58] The aim demonstrates the focus of NEMP is about what children "can do", which has the same function as Colwell's achievement test.

However, in *Learning Sequences in Music*, which Gordon investigates on how learning theory is applied to music, the comparative nature of developmental and stabilised music aptitudes, and the types and stages of music audiation, he points out the weakness in using achievement tests, he argues that "for teaching appropriately to students' individual musical differences, music aptitude scores are mandatory. When music achievement scores are used instead, they have much less value because they

---

[57] Colwell, R. (1969). *Music Achievement Tests.* Chicago: Follett Educational Corporation. 24.

[58] Ministry of Education, EARU. (2001) NEMP Music Assessment Results 2000. (*National Education Monitoring Report 17.).* Dunedin, NZ.: Crooks, T. and Flockton. L. p2.

can be deceptive."[59]

In Kirchhubel's thesis, we can find the answer to what Gordon meant by the scores being potentially deceptive. Gordon's more recent view on aptitude and achievement is that music aptitude is a measure of a student's potential or capacity to learn, whereas music achievement is by contrast a measure of what a student has learned. Kirchhubel believes Gordon's extensive researches on music aptitude and achievement "demonstrate(s) that students with high levels of music achievement also possess high levels of music aptitude, whereas students displaying low levels of music achievement, do not necessarily possess low levels of music aptitude."[60] The deception lies within the cases of students displaying low levels of music achievement because of learning or adaptation abilities rather than actual music aptitude. Gordon believes "Aptitude and achievement are not mutually exclusive."[61] "Music aptitude is a product of nature and nurture, both of which contribute in unknown proportions to a student's music aptitude."[62] Empirical evidence indicates that while students who score highly on a music achievement test, also tend to score highly on a music aptitude test, but students who score highly on a music aptitude test, might not necessarily score highly on a music achievement test, because the possession of a

---

[59] Gordon, E. E. (1980). *Learning sequences in Music: Skill, Content, and Patterns*. (1984 edi.). Chicago: IL: GIA. 223.

[60] Cited by Kirchhubel, J. (2002). *Adolescent Music Development and the Influence of Pre-Tertiary Specialised Music Training.* Unpublished doctoral dissertation. Griffith University, Brisbane, QSL, Australia. 26-27.

[61] Gordon, E. E. (1980). *Learning sequences in Music: Skill, Content, and Patterns*. (1984 edi.). Chicago: IL: GIA. 223.

[62] Ibid. 225.

refined sensory capacity does not imply high levels of music achievement, which may be attained through later training.

These differing adopted test names displays a non-mutually exclusiveness of the nature of ability measuring tests. Each researcher attempts to reflect precisely which view s/he holds with these different names, on whether s/he believes aptitude is relevant to whether cognitive development can be accumulative through age, or whether learning and aptitude are mutually exclusive or not. It also attempt to demonstrate what elements is s/he trying to measure, whether s/he thinks musical aptitude is defined by one's physical, psychological or neurological characteristics. This disagreement by researchers on a definite criterion for musicality highlights the basic problem of the validity of all musical ability tests.

Furthermore, to exemplify the existing disagreement by researchers on a definite criterion for musicality, the words such as 'ability', 'achievement', 'musicality', and 'talent' are sometimes used without defining them adequately, or even worse, used wrongly, and thereby confuse, aptitude and achievement.

In Gordon's view, aptitude is a measure of a student's potential to learn, and achievement is a measure of what he has learned. Music aptitude tests focus on the measurement of sensory factors, such as, pitch, loudness, rhythm, time, timbre and

tonal memory, whereas music achievement tests measure performance and non-performance skills, such as musical concepts, notation, and terminology.

We have discussed a few different test designs in this chapter. While the disagreement in the classification of all musical ability tests highlights a validity problem, it is unavoidable that "differing views on the nature of musicality must inevitably lead to different test designs and results"[63], thus providing that further investigation in musical ability tests are conducted with precision in measurement objectives, researchers and education theorists can agree to disagree.

---

[63] Buckton, R. M. (1981). *The Development of Musical Concepts in Young Children: an Investigation.* Unpublished doctoral dissertation. University of Auckland, Auckland, New Zealand. 29.

2.2 Measurement of attitude

Little effort has been made to develop attitude scales in music before the 1980's. "Most studies that measure attitude employ the interview technique or an arbitrarily devised questionnaire."[64] No statistically reliable instrument to measure attitudes has been developed. Consequently, current attitude research in music can have problem with establishing external validity.

Still, the complex nature of the measurement of attitude should not stop further research into the area. Mueller observed, "Since learning is so largely a matter of motivation and the desire to learn, perhaps the first tests should be concerned with these factors, with attitudes and values, and other personality traits involved in the will to learn."[65]

Jerry W. Morris and Michael H. Stuckhardt noted that, "art educators have used the term attitude too casually, without a specified meaning, and in contradictory ways. This minimum of congruent meaning and lack of definition for the term attitude has led to misuses, misunderstandings, and difficulties in the practical application of this concept to teaching and research practices."[66]

Morris and Stuckhardt defined attitude as follows:

---

[64] Schneider, E. H. & Cady, H. L. (1965). Evaluation and Synthesis of Research Studies Relating to Music Education. *Cooperative Research Project. E-016*, Office of Education, U.S. Department of Health, Education and Welfare. Columbus, Ohio: The Ohio State University Research Foundation, p.110.

[65] Mueller, K. H. (1956). Studies in Music Appreciation. *Journal of Research in Music Education*, *4*, 4.

[66] Morris, J. W., & Stuckhardt, M. H. (1977). Art Attitude: Conceptualisation and Implication. *Studies in Art Education. 19*(1), 21.

a) Attitudes are effective evaluative concepts which give rise to motivational behaviour (Allport, 1935; Bem, 1970; Fishbein, 1967a; Katz & Stotland, 1959; Krech, Crutchfield & Ballachey, 1962; Rosenberg, 1956);
b) Attitudes are learned (Allport, 1935; English & English, 1958, p. 50; Fishbein, 1967a, p. 483; Lemon, 1973, p.15; McGuire, 1969, p.147);
c) Attitudes have specific social referents (Newcomb, Turner & Converse, 1965; Sherif & Sherif, 1956);
d) Attitudes are relatively stable and enduring (Campbell, 1950, p. 51; Krech, et al., 1962, p.178; Newcomb, 1966, p.22; Rosenberg, 1956, p. 370);
e) Attitudes vary in quality and intensity (Krech, et all., 1962; McGrath, 1964);
f) Attitudes are interrelated (Krech, et al., 1962, p. 178; Shaw & Wright, 1967, p. 9; Williams, 1960, p.468).[67]

After defining the meaning of attitude, the question of which method is appropriate for measuring attitude arises. There are varying views on this subject, similar to previous music ability tests' situation. Leon Crickmore[68] stressed that the enjoyment of music is a mental act relatively independent of experimental studies involving with the measurement of aural abilities or various reactions and moods evoked by short musical extracts. On the other hand, McGuire (1969)[69] argued that direct contact with art is the major determinant of attitudes held toward art, that it is not possible to separate a learning experience from personal attitude, therefore attitude and learning come side by side. Both views have their own logical reasons supporting them, to find out which view is true, tests must be run to determine whether there is a relationship between attitude and learning.

---

[67] The quoted references have not been referenced here as they are part of the quotation of Morris, J. W., & Stuckhardt, M. H. (1977). Art Attitude: Conceptualisation and Implication. *Studies in Art Education. 19*(1), 21.

[68] Crickmore, L. (1968). An Approach to the Measurement of Music Appreciation. *Journal of Research in Music Education*, *16*(3), 239.

[69] McGuire, W. J. (1969). *The nature of attitude and attitude change.* In Linzey, G. & Aronson, E. (Eds.), The handbook of social psychology (Vol. 3). Reading, Mass.: Addison-Wesley. 166.

Before trying to find whether there is a relationship between the two, it is necessary to define the two subjects: attitude and learning. Attitude, being an aesthetic aspect makes it hard to measure. Thurstone believed it is necessary to make some assumption before proceeding to the measurement of attitude:

1. We take for granted that people's attitudes are subject to change. Using standard error of measurement to accommodate for the fluctuation in attitude.
2. Attitude scales will be used only in those situations that offer a minimum of pressure on the attitude to be measured.
3. Minimise as far as possible the conditions that prevent our subject from telling the truth, or else to adjust our interpretations accordingly.[70]

Based on the belief that 'Attitudes are learned'[71], Morris and Stukhardt suggested that,

> New attitudes can be taught and previously held attitudes can be reinforced, altered, or broken down through directed teaching….The significance of these attitude characteristics should not be overlooked by art educators. Students come to the art classroom with previously acquired attitudes toward art. Not only will students resist changes to these held attitudes, but they will actually seek to reinforce them.[72]

Still, attitude can be difficult to identify. For example, a teacher might think a student holds a negative attitude toward listening to music when in fact the attitude detected is held toward the environment or genre or style of music.

---

[70] Thurstone, L. L. (1928). Attitudes Can Be Measured. *The American Journal of Sociology*, *33*(4), 529-554.

[71] Morris, J. W., & Stuckhardt, M. H. (1977). Art Attitude: Conceptualisation and Implication. *Studies in Art Education. 19*(1), 23.

[72] Ibid.

Taking in the complex nature of attitude, and its importance to art education, Mittler noted that "specific provisions must be made within the curriculum for the learning of attitudes, and specific experiences or learning encounters must be developed which will effectively influence, foster, or inculcate desired attitudes."[73] Morris and Stukhardt also stressed this influence of attitude on learning should be explored and researched with more depth by art educators. Stukhardt reported that the few studies which have been conducted in the area of art attitudes are limited in their usefulness because of conceptual and procedural faults.[74] Paul Lehman also noted in "*Tests and Measurement in Music"* that, "although music aptitude and achievement have been subject to considerable investigation, little research has been conducted in attitude formation."[75] For these reasons, it is of value to look at research involved with other subjects relating to the area of attitude.

Attitude researches in other subjects

Japanese research using 'cross-lagged correlation', measuring the relationship between attitude and achievement in mathematics education found that attitude has greater effect on achievement, than the reverse.[76]

The importance of attitude on achievement is also apparent in study on rote learning.[77]

---

[73] Mittler, G. (1972). Efforts to secure congruent and incongruent modifications of attitude toward works of art. *Studies in Art Education, 13*(2), 58-70.

[74] Stuckhardt, M. H. (1973). A critical review of attitude scales used in the visual arts. *Review of Research in Visual and Environmental Education. 1*(2), 7-23.

[75] Lehman, P. (1968). *Tests and Measurements in Music.* Englewood Cliffs, N. J.: Prentice-Hall. 79-80.

[76] Minato, S., & Kamada, T. (1996). Results of Research Studies on Causal Predominance between Achievement and Attitude in Junior High School Mathematics of Japan. *Journal for Research in Mathematics Education*, *27*(1), 96-99.

[77] Tinkham, T. (1989). Rote Learning, Attitudes, and Abilities: A Comparison of Japanese and American Student. *TESOL Quarterly*, *23*(4). 695-698.

A comparison of the attitudes of Japanese and American students toward both rote and creative learning, a comparison of the same students' performances, given similar rote learning tasks was found to have the expected results. Japanese students with more positive score towards rote learning performed significantly better in both recalling and recognizing new words in a new language.

In the findings of an art study by Alison King, using one-way analysis of variance (ANOVA) test, she suggested that the process of freely choosing learning experiences may foster student achievement and enhance student self-concept and improve student attitude toward the subject. It appears that choice is a factor that can influence both achievement and attitude outcomes simultaneously.[78]

Up to this point, from the researches discussed, they seem to suggest that there is a relationship between attitude and achievement, and the relationship is present as attitude affecting achievement, rather than vice versa.

In Steven K. Hedden's research also showed a similar result. He asked the question: "What is the magnitude of the relationship between music achievement of fifth and sixth graders and a set of predictor variables – academic achievement, attitude toward music, self-concept in music, music background, and gender?" [79] In this study, it was found a teacher might be able to improve music achievement by stressing the development of positive attitudes towards music and positive self-concepts in music; both variables were significant predictors of music achievement. His results will be

---

[78] King, A. (1983). Agency, Achievement, and Self-Concept of Young Adolescent Art Students. *Studies in Art Education*, *24*(3), 187-194.

[79] Hedden, S. K. (1982). Prediction of Music Achievement in the Elementary School. *Journal of Research in Music Education, 30*(1), 61-68.

discussed further along with Cary's[80] finding in chapter five.

Gordon decided to approach the issue involved in understanding the relationship between attitude and learning/achievement differently. Instead of measuring attitude and learning as two independent elements, Gordon avoids the debate of whether the two are independent or not completely, by looking at the two from a different angle – to appreciate or to understand. "To appreciate means to evaluate highly or approve warmly often with expressions or tokens of liking. To understand means to grasp the meaning of." From the definition Gordon has provided, we can see in Gordon's view, attitude comes naturally in a learning situation, where the two cannot be looked at separated. He also pointed out that,

> Much of what is learned is not taught. As a result, what is taught is of great importance, because it provides the basis for what will be learned. Appreciation, like an understanding of music, can be developed only in stages. When a student is instructed according to a logical learning sequence, he learns also how to continually be his own teacher.[81]

Gordon noted,

> It would seem that appreciation suggests a favourable emotional response and understanding is based upon comprehension…. Understanding as the primary purpose of music education is gaining acceptance, and appreciation is being placed in its proper perspective. Understanding is the basis for appreciation; all but possibly the purely emotional reactions to music depend upon understanding. A student who appreciates music without understanding it demonstrates an emotional reaction, whereas a student who appreciates music and understands it demonstrates an aesthetic reaction. For a student to say that he appreciates a

---

[80] Cary, S. E. (1981). *Individualised Music Instruction – Traditional Music Instruction: Relationships of Music Achievement, Music Performance, Music attitude, Music Aptitude, and Reading in Classes of Fifth Grade Students.* Unpublished doctoral dissertation. Oregon University, Michigan.

[81] Gordon, E. E. (1980). *Learning sequences in Music: Skill, Content, and Patterns*. (1984 edi.). Chicago: IL: GIA. 2.

> piece of music without understanding that music is to admit that he is prejudice.… When taught to perform music with understanding, a student learns music. When taught to appreciate music, a student develops vague positive and negative attitudes, and simply learns about music.[82]

From the above quote, it demonstrates how Gordon's view on musicality test is the gestalt view approach, that musical ability and experience should be viewed together in a big picture where individual musical elements are not analysed separately.

In chapter five, this research will attempt to test the relationships between some of these elements, such as the relationship between attitude, aptitude and learning. It was necessary to discuss many representative theorists differing views on musicality and attitude to understand precisely what the results in chapter five represent. In this chapter, we have discussed many different designs of aptitude and achievement tests to understand the different testing methods available, then the intricacy in the measurement of attitude. Now we must compare NEMP – the test used in this research with other national/state assessments. This should provide an understanding of NEMP at an international level; hence provide ground for basing this research on NEMP student scores and survey results.

---

[82] Gordon, E. E. (1980). *Learning sequences in Music: Skill, Content, and Patterns*. (1984 edi.). Chicago: IL: GIA. 1.

2.3 National assessment in New Zealand and other countries

In the beginning of this chapter, we briefly discussed the different views of assessment tests to understand the fundamental structure of NEMP assessments, it is now necessary to explain the uniqueness of NEMP on an international level, thus justifying the reason for using NEMP scores and survey results as the basis of this research.

New Zealand is similar to Scotland in having a higher proportion of small primary schools in relation to total population.[83] In 1991 the curriculum was restructured in some changes modelled on the curriculum and assessment changes in the late 1980's in England and Wales.[84] Thus, there are strands within subject areas and achievement objectives at eight levels within each strand. Still, New Zealand is among the few countries, which include music as a nationally assessed subject and conducts surveys to evaluate students' attitude. Most countries concern themselves with compulsory national assessments involving information on numeracy, oracy and written language. The subjects being assessed are therefore English, mathematics and science. A few countries with low-stake national assessments are discussed in this section of the chapter.

In Education for All (EFA)'s *Global Monitoring Report 2008,*[85] these following countries have music as a curriculum subject: New Zealand, Montenegro, Serbia,

---

[83] Harlen, W. (2007). The quality of learning: assessment alternative for primary education. *Primary Review Research Survey 3/4,* Cambridge: University of Cambridge Faculty of Education, 10.

[84] Crooks, T. J. (2002). Educational Assessment in New Zealand Schools, *Assessment in Education, 9*(2), 239.

[85] The report is based on national learning assessments undertaken between 1995 and 2006.

United States, United Kingdom (Wales). New Zealand, United States and United Kingdom (Wales) will be discussed with more details in this research. Because NEMP of New Zealand and NAEP of the United States were already discussed in chapter I, there will be more focus on Wales.

Assessment in New Zealand primary schools is predominantly low-stakes assessment.[86] NEMP focuses on monitoring pupils' learning, improving learning through direct feedback to students or adjustments to teaching programme.[87] Crooks believes the assessment results were used within schools for monitoring progress and standards and by the inspectorate[88] for external review, therefore "do not have a dominant influence on teachers' assessment practice."

The low-stakes is preserved by the existence of a quite separate programme for national monitoring. In the case of New Zealand, while using NCEA (National Certificate of Educational Achievement) as the standard national examination for secondary education, in primary education, NEMP was developed for a monitoring purpose, and the Scottish Survey of Achievement is used in Scotland for the same purpose. The biggest difference between the two lies in the fact that New Zealand treats music within the arts curriculum, but in Scotland, music is treated as a club activity in and after school and not a curriculum subject.

---

[86] Crooks, T. J. (2002). Educational Assessment in New Zealand Schools, *Assessment in Education, 9*(2), 246.

[87] Ibid.

[88] After 1989, The Education Review Office is responsible for auditing schools against member-level legislation, school charters, and other, policy requirements. An inspectorate used to be allocated with new teachers to assess and certify their competence. Today, new teachers are no longer certified, but registered.

In Scotland, Wales and Northern Ireland of the United Kingdom, there have been considerable changes in the systems of assessment. Scotland began its major reform in 1999, while Wales began to be in the process of phasing in change, and Northern Ireland having had its policy and organisational change in 2006.[89] Although these countries were at different points in implementation of change, there was a common direction of moving towards "greater use of assessment by teachers and away from frequent testing, to warrant discussing them together."[90]

The curriculum and assessment in place until 2000 were established by the same Education Reform Act of 1988 that applied to both England and Wales. In 2000, the Wales' *Curriculum 2000* was introduced and the decision was taken to end statutory tests and tasks at the end of Key Stage 1.[91] Statutory assessment by teachers was the only form of assessment at the end of Key Stage 1, whereas at the end of Key Stage 2, both teachers' assessment and results of tests were reported with equal importance.

ACCAC (then the Qualifications, Curriculum and Assessment Authority for Wales, now within the Department for Children, Education, Lifelong Learning and Skills (DCELLS) of the Welsh Assembly Government) recommended extensive changes in the assessment system (ACCAC, 2004). Teacher's judgements were monitored by a moderating system specifically set up to ensure acceptable accuracy and consistency. Teachers' assessment extended further from end of Key Stage 2 to 3. Because of this change, and to move teachers away from relying on test-derived data, centrally

---

[89] Harlen, W. (2007). The quality of learning: assessment alternative for primary education. *Primary Review Research Survey 3/4,* Cambridge: University of Cambridge Faculty of Education, 7.

[90] Ibid.

[91] A Key Stage is a stage of the state education system in England, Wales and Northern Ireland setting the educational knowledge expected of students at various ages. Key Stage 1: 5-7 years old, Key Stage 2: 7-11 years old, Key Stage 3: 11-14 years old, Key Stage 4: 14-16 years old.

produced guidance and support were given to schools in setting up procedures.

Previously in Chapter one, high-stakes and low-stakes testing were discussed. When considering assessments around a few different countries there is evidence that there is benefit in conducting low-stakes testing. In addition, changing teachers' assessment can encourage a richer curriculum experience for pupils. For example, Harlen (2007) noted that, Flexer et al (1995) reported changes when teachers of third grade pupils in a school district in the USA were introduced to assessment methods using evidence from pupils' classroom performance instead of using tests. Several effects were found on teachers and pupils after a year of using these methods:

"Teachers were using more hands-on activities, problem solving and actively asking pupils for explanations. They were also trying to use more systematic observations for assessment. All [teachers] agreed that the pupils had learned more" [92] than under the previous system and that they know more about what their pupils could achieve.

The work of Flexer et al (1995) shows that in a low-stakes testing environment, teachers use a much wider range of assessment methods.[93] Further evidence was provided by Hall and Harding (2002) and Hall et al (1997), who reported on the introduction of teachers' assessment in the National Curriculum[94]

---

[92] Harlen, W. (2007). The quality of learning: assessment alternative for primary education. *Primary Review Research Survey 3/4,* Cambridge: University of Cambridge Faculty of Education, 20.

[93] Flexer, R.J., Cumbo, K., Borko, H., Mayfield, V and Maion, S.F. (1995). How 'messing about' with performance assessment in mathematics affects what happens in classrooms. *Technical Report, 396, Los Angeles Centre for Research on Evaluation, Standards and Student Testing (CRESST).*

[94] Hall, K. & Harding, A. (2002). Level descriptions and teacher assessment in England: towards a community of assessment practice. *Educational Research, 44*. 1-15.

Hall, K., Webber, B., Varley, S, Young, V & Dorman, P. (1997). A study of teachers' assessment at Key Stage 1. *Cambridge Journal of Education, 27.* 107-122.

Assessment used in England and Wales, was perceived by teachers as having a positive impact on pupils' learning. Harlen (2007) cited Hall and Harding (2002) that,

"Their summative assessment was based on teachers' judgements across a range of pupils' work. The impact was enhanced by teachers working collaboratively toward a shared understanding of the goals and procedures to achieve these goals. Unfortunately the funding and opportunities for these meetings declined due to pressure to raise test scores and the ground that was gained (in quality of teacher assessment) in the early and mid '90s was lost."[95]

This is a different situation from NAEP of USA which was discussed in Chapter one. NAEP through the years have opted for easier procedures and better public communications, such as replacing professionally trained administrator with local school personnel and giving more attention to certain subjects. These moves have raised questions about the uniformity of administration and credibility of representing population's progress in levels of educational achievement of USA.

France and Scotland have similar surveys like New Zealand for monitoring regional and national standards, but not in the subject of music. This method is valuable for schools assessed to focus attention on their own practice and the performance of pupils in the areas identified as weaknesses. It not only encourages school participation in the surveys, but decreases the possibility of adding high-stakes to the assessment of pupils. Scotland and France, and Wales, who is planning to join this scheme, avoid the use of high-stakes tests. The evaluation of teachers and schools surveyed in these countries are based on a range of indicators relating to the context,

---

[95] Cited by Harlen, W. (2007). The quality of learning: assessment alternative for primary education. *Primary Review Research Survey 3/4,* Cambridge: University of Cambridge Faculty of Education, 21. Based on Hall, K. & Harding, A. (2002). Level descriptions and teacher assessment in England: towards a community of assessment practice. *Educational Research, 44*. 13.

environment, curriculum provision and resources as well as pupil performance. New Zealand achieves a similar result by considering possible differences in performance patterns for different demographic groups and categories of school while still remaining anonymous.

Unlike New Zealand's low-stake assessment approach, similar to Wales before their planned change, Japan believes a national sample survey should provide for both diagnostic and informative purposes. It should provide the assessed pupils and schools "with tools to improve their practice rather than to be used by others to control teachers and schools."[96] The National Institute for Educational Research in Japan (NIER) conducted its first national survey of achievement since the 1960s in 2007. The purpose of the assessment is to respond to the growing concerns in declining academic abilities in children. Test results for each municipality and school were not announced to avoid compiled league tables, but local governments and schools received the individual test results and compared the students' position to a national average, and subsequently take steps of their own to improve academic abilities. Students also received their results so that they would know which areas to improve on.

Ireland's National Monitoring System is another example of low-stakes testing. It aims to identify changes in trends over time for specific administered areas of schools, depending on which area the school intends to find more information on, and to inform ongoing policy development while individual school results remain unidentifiable. Canada's PCAP assessments, like New Zealand, include a

---

[96] Harlen, W. (2007). The quality of learning: assessment alternative for primary education. *Primary Review Research Survey 3/4,* Cambridge: University of Cambridge Faculty of Education, 16.

questionnaire about students' interests, attitudes in classroom activities. Results of PCAP assessments, individual students, schools, or school board/school district are not reported. Students' academic records are not affected. Hungary's statutory assessment is not a high-stakes test either. It aims to gain a general picture of the effectiveness of education and of trends in student achievement.

Australia is the closest neighbouring country, unlike New Zealand, it does not have a national or state monitoring assessment for music. Similar to many other countries, it concentrates on literacy and numeracy.[97]

Many education systems were discussed and compared in this chapter. The aim was not to declare which method(s) might be superior, but to identify and compare New Zealand's NEMP being a justifiable system on an international level.

---

[97] Andrews, C. Brown, R. et al., (2007). Compulsory assessment systems in the INCA countries: Thematic Probe, *INCA Internet Archive*. 4.

2.4 Summary

In this research, the aim is to find if there is a relationship between attitude and ability by using results from the year 2000 NEMP scores and survey results. Earlier this chapter, different ability-measuring tests were discussed. Gestalt, atomist and other views were mentioned. Some focused on innate ability, some on learned skills and knowledge, and NEMP lies in the middle, measuring students' abilities without separating aptitude and achievement. Background knowledge regarding the recent literature of testing in the discipline of music can assist understanding the positioning of NEMP therein.

Much literature demonstrates the problem of measuring attitude. Problems exist when trying to measure differences in opinions.[98] Different researchers have tackled this problem differently with varying results. Most discussed within the grounds of attitude and learning, while Gordon looked at the two from a different angle – to appreciate and to understand. Assumptions were being made for attitude scales. Since attitude is being used in this research, it is necessary to recognise the limitations of analysing attitude because of the influence it would have on the results of this research.

NEMP scores and survey results are the basis of this research. It was chosen for its uniqueness at an international level. To explain the uniqueness of NEMP, education

---

[98] Thurstone, L. L. (1928). Attitudes Can Be Measured. *The American Journal of Sociology*, *33*(4), 532.

systems and tests of different countries were discussed and compared. In Education for All (EFA) Global Monitoring Report 2008, they provided a global overview of national learning assessments undertaken between 1995 and 2006. Out of all the countries, there were only five which has music as a curriculum assessed subject at a national level, and New Zealand is one. The predominantly low-stakes assessment in New Zealand, Wales and many others, focuses on monitoring pupils' learning, improving learning through direct feedback to students or adjustments to teaching programme. It would be interesting to see whether New Zealand will change in the way that NAEP of USA, that is further away from low-stakes assessment or otherwise. Although France and Scotland have similar surveys to New Zealand, they are not used in music. Thus, NEMP provides an invaluable opportunity to study issues such as contained in this thesis.

Chapter III - Procedures

3.1 Introduction

This research aims to study the relationship between student attitude and performance as measured by selected musical tasks in the 2000 National Education Monitoring Project (NEMP) scores and survey results. Part of NEMP results were used to represent students' musical abilities. Still, there may be other factors which could potentially influence a student's results in the NEMP assessment, and these will be discussed in 4.6, the limitations section.

The null hypothesis of the research follows:

H_0: There is no relationship between a student's attitude towards music and his/her musical abilities as measured by the NEMP scores and survey results.

"A null hypothesis is a statement that asserts the status quo; that is, that any change from what has been thought to be true is due to random sampling error."[99] For the purpose of this study, parts of NEMP results were used with the assumption that they demonstrate students' musical abilities. Factors relating to student's musical experiences were also tested concerning the results to see how much other elements might have influenced students' musical abilities.

---

[99] Hair et al. (2003) *Marketing Research – Within a changing Information Environment (2nd ed.).* New York, NY: McGraw-Hill. 538.

3.2 Research methods and procedures - explanation of the relevance and importance of differing variables

A causal research design was used to test the hypothesis of whether a relationship exists between a student's attitude towards music and his/her musical abilities.

The secondary information[100] used in this study are 178 tapes and surveys of year 2000 obtained from the Educational Assessment Research Unit of the University of Otago, which conducts the NEMP programme.

Quantitative data was used in this research. Because the researcher had access to the video tapes, the individual scores were analysed by the researcher herself, using criteria apposite to the needs of this study, as distinct from those applicable to the NEMP programme.

The NEMP student samples were randomly selected from participating school around New Zealand using computer generated random numbers. 1440 year 4 and 1440 year 8 children respectively were chosen to represent about 2.5 percent of the children at those levels in New Zealand schools. Different students attempted different tasks. The 1440 students selected at each level were divided into three groups (A, B, C) of 480 students, comprising four students from each of 120 schools.

---

[100] Secondary information is information that was collected and interpreted at least once for some specific situation other than the current one. Cited in Hair et al (2003). 59-60.

Due to time constraint, 90 tapes and surveys (6.25% of the available sample) from each level of the tasks done by Group C were randomly selected for the use of this research. This randomly selected sample represents 0.14 percent of New Zealand children at each of year 4 and 8 levels. The Group C students completed four one-to-one tasks, all involved keyboard type of instruments: electrical keyboard or chime bars. These were chosen because these tasks involved both sound and rhythmic elements, and the one-to-one approach made it possible to analyse individual student.

Analysis of variance (ANOVA) and Pearson correlation analysis were used to examine differences between gender, students' attitudes from NEMP survey results, and experiences in relation to their NEMP scores.

The independent variables are musical activities, students' attitudes and experiences, and dependent variables are actual scores of sampled children.

The relationships between the variables of the children's scores and the survey results were tested with chi-square analysis. Chi-square analysis was used to assess the significance between the frequency distribution of male and female students. This result was used to compare with the sample published by NEMP. "The use of chi-square statistic is very helpful in answering questions about data that are nominally scaled and not suitable for other types of statistical analysis, such as ANOVA or t-tests."[101] The significance level at 0.05, 95% confidence interval which is used in most researches will be considered, because confidence interval at 100%

[101] Ibid. 566.

consist of extreme values which are best to be eliminated, and 90% does not possess sufficient accuracy level.[102]

Pearson correlation coefficient measures the strength of a linear relationship between two metric variables. It varies between -1.00 and +1.00, with 0 representing absolutely no association between two variables, and -1.00 or +1.00 representing a perfect link between two variables. The higher the correlation coefficient, the stronger the level of association.[103]

Table Two

Rules of Thumb about the Strength of Correlation Coefficients[104]

Range of Coefficient	Description of Strength
±.81 to ±1.00	Very Strong
±.61 to ±.80	Strong
±.41 to ±.60	Moderate
±.21 to ±.40	Weak
±.00 to ±.20	None

[102] Ibid. 539.
[103] Ibid. 567-568.

[104] Ibid. 568.

3.3 The main sample

The samples of 96 year 4 and 87 year 8 students were randomly selected by EARU. Because this is only 6.25% of the original 2.5% sample population, the average of each question was compared to assess the reliability of the sample used in this research.

The sample of the student responses used in this research was compared with the original NEMP sample:

Original NEMP Sample

Question/instructions:		% responses y4	y8
In this activity we're going to use the key-board.			
1. Do you play the piano or keyboard at home or at school?	yes	37	41
2. Have you had lessons on the piano or keyboard?	yes	20	31

On p27 of year 2000 NEMP assessment, the NEMP sample shows 37% year 4 and 41% year 8 students play the piano or keyboard at home or at school, which is similar to the sample this research has used (Table 3, p67), 33.3% year 4 and 40.2% year 8 students play the piano or keyboard at home or at school.

20% year 4 and 31% year 8 students have had lessons on the piano or keyboard in the NEMP sample, while 20.8% year 4 and 29.9% year 8 students have had lessons on the piano or keyboard in the research sample.

Up to this point, the sample results are similar.

There is greater than 5% difference in the correct second attempts for Year 4's questions one, whereas for Year 8, there is more than 5% difference are present in 1st correct attempts of questions one, two, five and six, and 2nd correct attempts of question one. In most cases, except in the first correct attempts of Year 8 in questions five and six have sample percentages more than 5% less than NEMP percentage.

Table Three

Sample comparison for Task One by year

P27 Questions	Correct Attempts	NEMP Year 4 / Sample Year 4 (%)	NEMP Year 8 / Sample Year 8 (%)
1	1st	59 / 62.5	79 / 60.9
	2nd	17 / 10.5	12 / 5.7
2	1st	20 / 22	50 / 42.5
	2nd	12 / 7.3	10 / 11.5
3	1st	19 / 22.9	56 / 55.2
	2nd	7 / 6.2	10 / 11.5
4	1st	25 / 26	50 / 54
	2nd	11 / 9.4	17 / 12.5
5	1st	3 / 1	4 / 9.1
	2nd	3 / 2	8 / 4.6
6	1st	2 / 1	4 / 10.3
	2nd	1 / 2.1	6 / 4.6

Results from Task Two were not compared due to the significant disparity in interpreting student responses between the original NEMP data and the method used by the researcher.

The percentage of survey results between the original NEMP sample and samples used in this research were reasonably similar. The percentage used in this research had taken the invalid cases into calculation, whereas the published NEMP sample results do not include invalid cases. It was not possible to measure the significance of variance without the original NEMP sample data, therefore, a list of different percentages are given without an indication of whether the differences are significant or not.

Table Four

Sample comparison for students' attitudes/experiences by year

Year 2000	NEMP sample Year 8 (%)	Research sample Year 8 (%)	NEMP sample Year 4 (%)	Research sample Year 4 (%)
How much do you like doing music at school?				
Very much	33	36.8	57	49
Reasonably	48	41.4	31	30.2
Okay	15	13.8	8	8.3
Not at all	4	1.1	4	1
How much do you do these things in music at school?				
Singing				
Lots	16	10.3	31	30.2
Quite often	34	37.9	32	29.2
Sometimes	44	41.4	35	32.3
Never	6	5.7	2	2.1
Playing instruments				
Lots	14	11.5	12	13.5

Year 2000	NEMP sample Year 8 (%)	Research sample Year 8 (%)	NEMP sample Year 4 (%)	Research sample Year 4 (%)
Playing instruments				
Quite often	28	29.9	18	16.7
Sometimes	46	47.1	58	55.2
Never	12	6.9	12	10.4
Listening to music				
Lots	28	29.9	33	25
Quite often	34	35.6	31	33.3
Sometimes	35	24.1	32	31.3
Never	3	5.7	4	4.2
Dancing/moving to music				
Lots	11	8	19	13.5
Quite often	19	21.8	20	19.8
Sometimes	51	52.9	45	44.8
Never	19	12.6	16	15.6
Making up music				
Lots	8	4.6	11	12.5
Quite often	16	19.5	15	10.4
Sometimes	48	43.7	36	35.4
Never	28	27.6	38	35.4
How much do you like doing these things in music at school?				
Singing				
Lots	26	23	52	52.1
Quite often	37	36.8	30	30.2
Sometimes	25	26.4	14	8.3
Never	12	10.3	4	5.2
Playing instruments				
Lots	47	37.9	65	61.6
Quite often	35	44.8	25	25
Sometimes	14	10.3	6	4.2
Never	4	2.3	4	2.1
Listening to music				
Lots	70	66.7	57	57.3
Quite often	21	21.8	32	32.3
Sometimes	7	5.7	8	4.2

Year 2000	NEMP sample Year 8 (%)	Research sample Year 8 (%)	NEMP sample Year 4 (%)	Research sample Year 4 (%)
Listening to music				
Never	2	1.1	3	2.1
Dancing/moving to music				
Lots	33	28.7	43	40.6
Quite often	30	31	28	19.8
Sometimes	24	23	16	19.8
Never	13	12.6	13	15.6
Making up music				
Lots	25	26.4	39	30.2
Quite often	37	33.3	27	29.2
Sometimes	27	26.4	18	19.8
Never	11	9.2	16	15.6
How much time out of school do you do these things in music?				
Singing				
Lots	18	21.8	30	28.1
Quite often	17	14.9	19	17.7
Sometimes	42	39.1	32	32.3
Never	23	19.5	19	16.7
Playing instruments				
Lots	16	13.8	20	20.8
Quite often	17	10.3	15	8.3
Sometimes	37	46	35	37.5
Never	30	24.1	30	28.1
Listening to music				
Lots	69	67.8	51	49
Quite often	20	18.4	25	28.1
Sometimes	8	8	19	14.6
Never	3	0	5	3.1
Dancing/moving to music				
Lots	19	19.5	23	22.9
Quite often	18	19.5	18	15.6
Sometimes	39	36.8	38	32.3
Never	24	18.4	21	24

Year 2000	NEMP sample Year 8 (%)	Research sample Year 8 (%)	NEMP sample Year 4 (%)	Research sample Year 4 (%)
Making up music				
Lots	8	6.9	17	14.6
Quite often	10	13.8	14	15.6
Sometimes	38	33.3	32	32.3
Never	44	39.1	37	32.3
How much do you like doing these things out of school time?				
Singing				
Lots	33	32.2	50	42.7
Quite often	30	32.2	23	24
Sometimes	21	17.2	15	14.6
Never	16	14.9	12	14.6
Playing instruments				
Lots	34	31	49	42.7
Quite often	35	39.1	27	32.3
Sometimes	17	14.9	11	11.5
Never	14	10.3	13	9.4
Listening to music				
Lots	89	88.5	71	94.8
Quite often	9	4.6	20	64.6
Sometimes	2	2.3	7	25
Never	0	0	2	5.2
Dancing/moving to music				
Lots	36	36.8	44	38.5
Quite often	26	19.5	23	26
Sometimes	23	27.6	17	13.5
Never	15	11.5	16	17.7
Making up music				
Lots	21	17.2	36	31.3
Quite often	26	34.5	23	20.8
Sometimes	28	23	18	21.9
Never	25	20.7	23	20.8

Chapter IV – Presentation and interpretation of results

Four tasks were used:

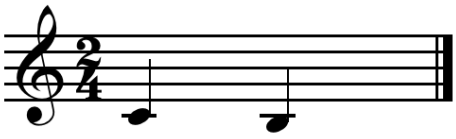
Task One p27 (2000) Year 4 and Year 8[105]

Coloured Keys

Students are expected to listen to notes played on a keyboard (always starting on middle C), and then repeat the pattern on the keyboard:

Q1

Q2



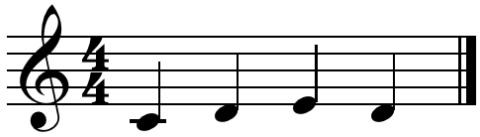
For statistical analysis, the following are the score guidelines used by the researcher:

P27Q1-2

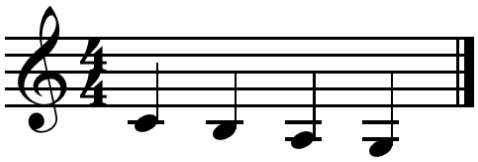
Score	Description
1	Wrong both times wrong pitch direction
2	Starting on wrong note wrong pitch direction one or both times
3	1st wrong 2nd same note
4	Wrong both times right pitch direction
5	Starting on wrong note correct pitch direction both times
6	1st no attempt 2nd correct
	1st correct pitch direction 2nd correct
	1st wrong pitch direction 2nd correct
7	1st correct 2nd wrong
8	Correct both times

[105] The original exercise with instructions can be found in appendix in a separate volume.

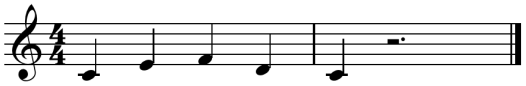
Q3



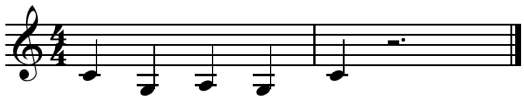
Q4



Q5



Q6



P27Q3-6

Score	Description
1	Wrong both times wrong shape
2	Starting on wrong note wrong shape one or both times
3	N/A
4	Wrong both times right shape
5	Starting on wrong note correct shape both times
6	1st no attempt 2nd correct
	1st correct shape 2nd Correct
	1st wrong shape 2nd Correct
7	1st correct 2nd wrong
8	Correct both times

The original NEMP scores only present students answering correctly at first or second attempt. The researcher listed more detailed scoring guideline to try to identify the possible reasons for students who answered incorrectly, and by grouping scores into different levels, to compare with different levels of students' attitude.

On the instructions for this question, teacher administrators were instructed to "ensure student locates middle C when instructed to play" the excerpts, and remind the student if s/he does not attempt each item starting on middle C. Still, there were many cases where students did not start on middle C and the administrators failed to inform the

students.

Students were asked if they played the piano or keyboard at home or at school, as well as whether they have had lessons on the piano or keyboard. They were also asked to play high and low notes on the piano, and demonstrate if they were capable of identifying high and low notes.

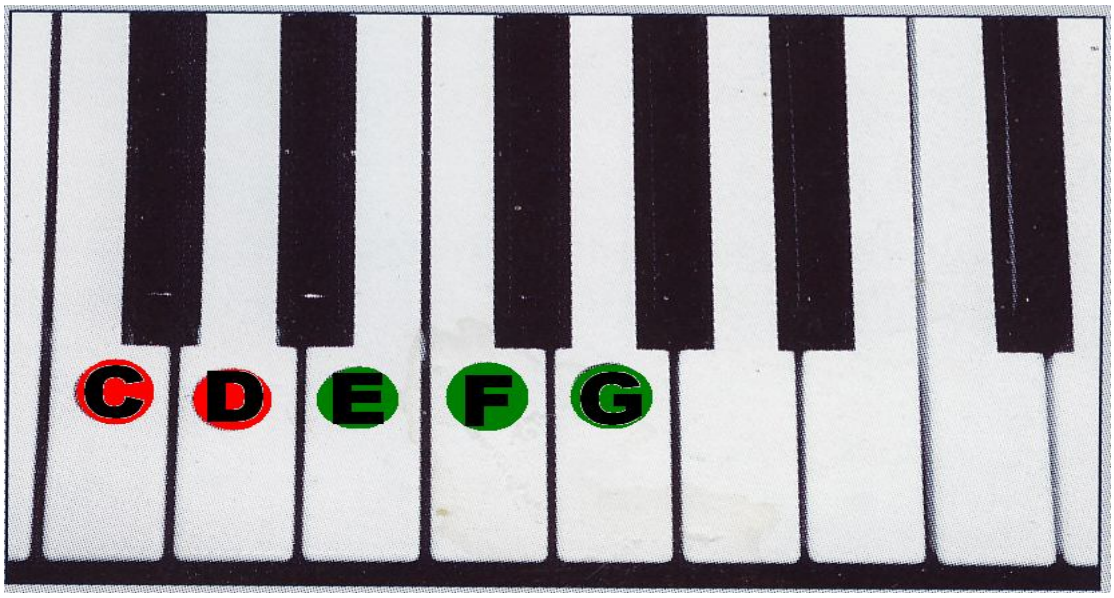
Task Two p22, 23 (2000) for Year 4

Keyboard Rhythms

In this task, the focus is to play and maintain note patterns rhythmically as an ostinato on an electronic keyboard, maintaining them against a melody, and in the last question, students are to invent a melodic pattern to fit rhythmically with the given melody.

This is an exercise which involves establishing a rhythmic beat and maintaining a melodic ostinato which will fit with the given music.

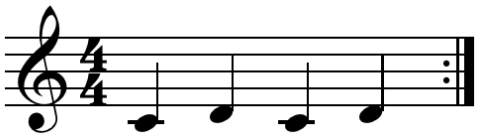
It is more important to establish a rhythmic pattern rather than a melodic pattern, because the original NEMP gave students scores for establishing a pattern and different levels of which the pattern fits with the melody, rather than student's selection of notes against the melody. In questions 1 to 4, students are given the pattern to maintain, but in question five students are required to create a pattern using the green notes.



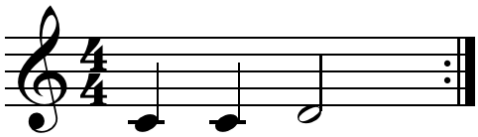
Students are told in questions 1 and 2 to use the red notes, and in questions 3, 4 and 5 to use the green notes. The stickers are labelled CDEFG.

Students are asked to play the given melodic patterns along with melodies heard on a video. They are asked not to watch the video but to listen to it and look at their hands on the keyboard.

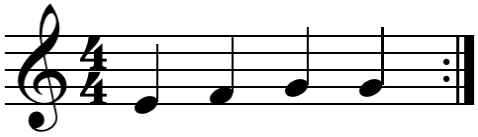
Q1



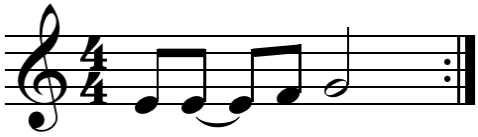
Q2



Q3



Q4



The original responses were categorised as follows:

Maintain pattern with melody:

1. throughout or most of time
2. late arrival at pattern
3. early loss of pattern
4. pattern not achieved

Researcher wanted to distinguish different levels of pattern establishment, the student responses were categorised into the following:

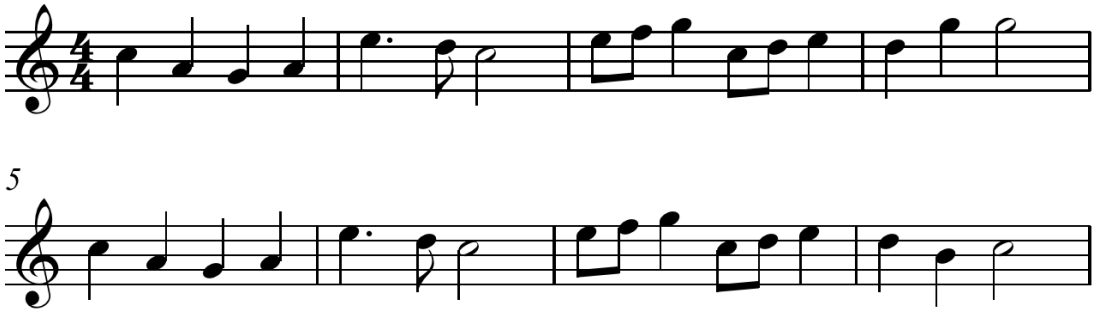
1. pattern not achieved
2. pattern not achieved and pattern attempted not in sync with the melody
3. pattern achieved but not in sync with the melody
4. late arrival of pattern / early loss of pattern but in time with the melody
5. pattern achieved half way with some missing notes or beats
6. pattern achieved and is in sync with the melody most of the time

7. pattern achieved with occasional mistakes
8. pattern achieved with correct speed

For the purposes of analysis, students were given a score from 8 to 1, 8 given to students who are able to maintain a pattern with the video's pattern throughout, and 1 given to students who were not able to establish a pattern.

In question 5, students are asked to make up their own melodic pattern using any of the green dots while the teacher on the video played the tune.

Q5



The original responses are categorised into:

Create pattern to fit with this melody.

Appropriate rhythm and timing:

1. none of the time
2. some of the time
3. all of the time

Choice of pattern:

1. never established
2. simpler
3. complex

A different scoring method was used in this research, credit was given for the following points, each student can score from 0 to 5, 0 for students not achieving any of the following points, and 5 for achieving all points:

- pattern established (a grouping of 1/2/4 is present)
- able to maintain with music
- pattern stays consistent
- use of complex pitch (chords or more than three pitches)
- use of complex rhythm (dotted rhythm or rests or note values longer than crotchet)

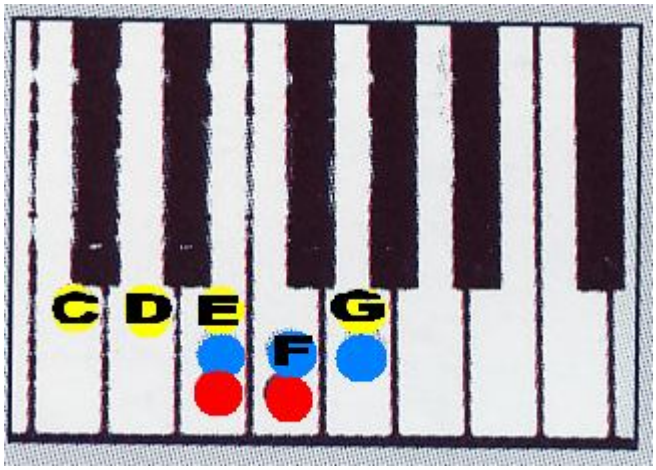
In the commentary section of this task, it was pointed out that "it would have been more appropriate to use the term "keyboard pattern" in the instructions for this task, as in the year 8 version". The original name "Keyboard Rhythms" failed to recognise the melodic pattern aspects of the task.

Task Two p24 (2000) for Year 8

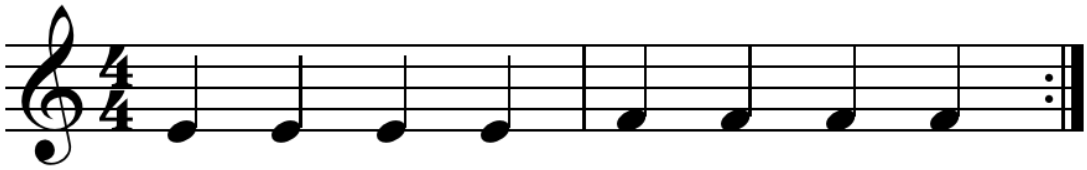
Keyboard Patterns

This exercise is similar to year 4 p22, 23 Keyboard Rhythm task. Students are asked to play melodic patterns rhythmically on an electronic keyboard, maintaining the patterns against a melody, and in the final question, inventing a melodic pattern with steady rhythm to fit with a given tune.

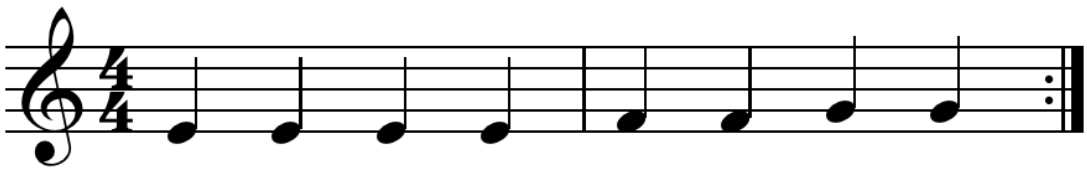
For their melodic patterns, students are asked to select notes from the two red notes for question 1, three blue dots for question 2, and four yellow dots for question 3.



Q1



Q2

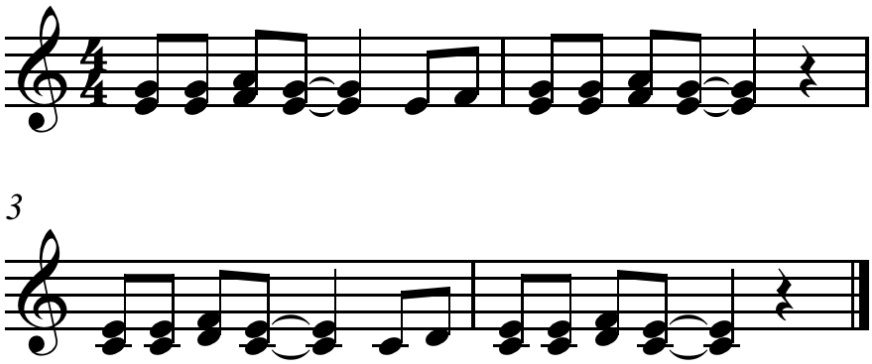


As in the year 4 task, a different scoring guideline was adopted. For questions one and two, student responses were categorised into the following:

1. pattern not achieved
2. pattern not achieved and pattern attempted not in sync with the melody
3. pattern achieved but not in sync with the melody
4. late arrival of pattern / early loss of pattern
5. pattern achieved half way with some missing notes or beats
6. pattern achieved and is in sync with the melody most of the time
7. pattern achieved with occasional mistakes
8. pattern achieved with correct speed

Students were given a score from 8 to 1 in this question, 8 given to students who were able to maintain a consistent pattern with the melody, and 1 given to students who were not able to establish a pattern.

Q3



Similar to year 4 p23 q5, a different score was given in this research, credit is given for the following points, each student can score from 0 to 5:

- pattern established (a grouping of 1/2/4 is present)
- able to maintain with music
- pattern stays consistent
- use of complex pitch (chords or more than three pitches)
- use of complex rhythm (dotted rhythm or rests or note values longer than crotchet)

Task Three p19 Link Test One (2000) / p13 (2004) Year 4 and Year 8

Fun Day

In this task, students are required to compose music for an advertisement using chime bars.

Lyrics were laid out as follow:

Monday's
Fun day
Down at the school
Fun day's
Monday
Come to the school

Student responses are categorised by NEMP as follows:

Rhythmic pattern fits words:

1. notes rarely matching syllables
2. a note matching most syllable
3. a note matching each syllables
4. a note matching each syllable plus a regular pulse

Presenting a lively well-phrased message: (consider vocal performance and tune)

1. not effective
2. only one element (voice or tune) present
3. moderately effective (both elements)
4. very effective (both elements)

A different scoring system was used in this research. Each point was given for achieving the following, 4 for achieving all points and 0 for none:

- Note matching each syllable
- In tune
- Sung rather than spoken for the vocal performance
- Regular pulse throughout

Task Four p28 Link Test Four (2000) / p20, 21 (2004) Year 4 and Year 8

Play It

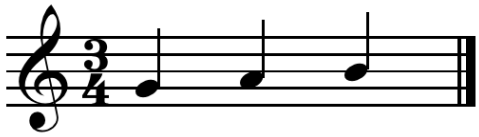
Students are asked to choose between using the keyboard or the chime bars for this task. They are asked to demonstrate a piece or two on the keyboard/chime bars.

The responses to their performance on the pieces were categorised into:

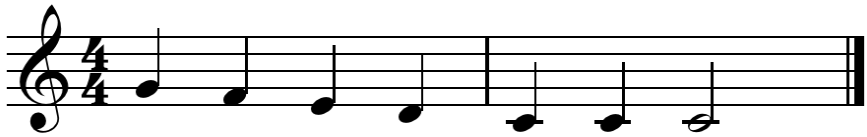
0. no piece played
1. piece played poorly
2. simple piece, played moderately well
3. complex piece, played moderately well
4. simple piece, played well
5. complex piece, played well (e.g. 2 hands playing).

Students are told the starting note is G and are told of its position on the keyboard/chime bars. If a student indicates that s/he cannot play the tunes, teacher administrators were instructed to discontinue the task.

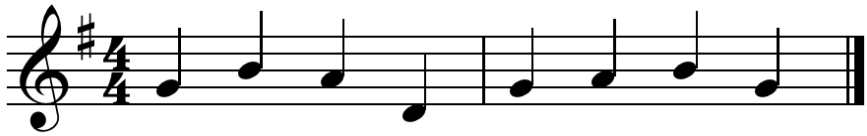
Q1



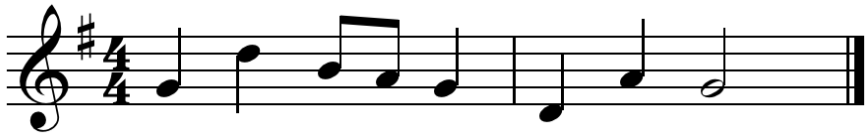
Q2



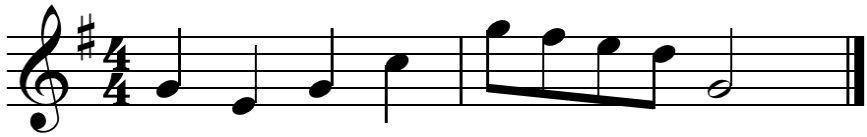
Q3



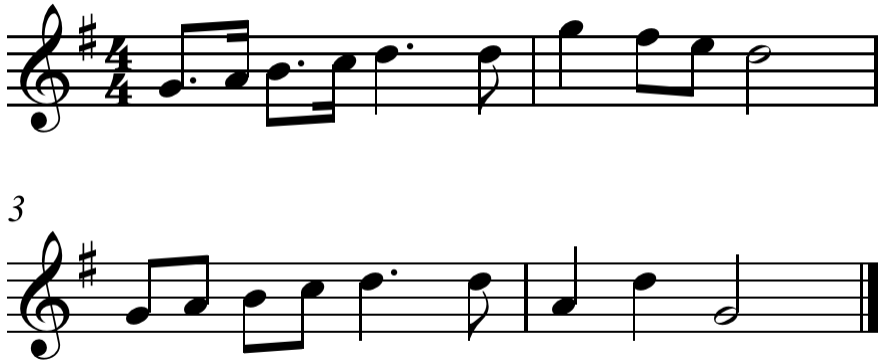
Q4



Q5



Q6



Students' responses are separated between note sight-reading ability, and rhythm/timing sight-reading ability.

A different scoring was used in this research to include the recognition of more advanced musical knowledge such as the recognition of key signatures and more complex note values.

Sight-reading: Notes

1. no attempt
2. limited attempt
3. half or fewer notes correct
4. mostly correct

5. correct throughout except recognising key signature F sharp
6. correct throughout

Sight-reading: Rhythm/Timing

1. no attempt
2. inaccurate
3. able to play with a steady crotchet beat
4. recognise quavers and minims and is mostly accurate
5. recognise dotted rhythm and almost all accurate
6. correct throughout

Note. Students who choose chime bars can only play the first three questions because of the note range of the chime bars. The chime bars consist of notes from middle C to the C an octave above.

4.1 Results of gender analysis

Key Assumptions

Although the NEMP tests were administered and marked by different people at a number of different locations, it is assumed that these factors do not greatly affect the results; and that furthermore, as the performances in the sample tapes will be assessed by the researcher's criteria, consistency will be maintained. Some answers were analysed with greater details. For example, rather than correct or incorrect scores, an incorrect attempt was dissected into smaller categories of pitch, rhythm and direction. More specified marking criterion were created by the researcher to illustrate the links.

T-test is a hypothesis test that utilizes the t distribution; used when the sample size is smaller than 30 and the standard deviation is unknown. T-test is more in common than the z-test. Both tests can be used for larger samples and known population standard deviations. T-test is used in this research. The population in this case is male=93, female=85.

Table Five

Mean of students' scores for NEMP tasks by gender

Group Statistics	Gender	N	Mean	Std. Deviation	Std. Error Mean
Overall score p27q1	Female	85	6.95	2.165	.235
	Male	93	6.62	2.340	.243
Overall score p27q2	Female	85	4.45	3.293	.357
	Male	93	3.81	2.939	.305
Overall score p27q3	Female	85	5.19	2.922	.317
	Male	93	4.92	2.589	.268
Overall score p27q4	Female	85	4.89	3.240	.351
	Male	93	4.11	3.315	.344
Overall score p27q5	Female	85	2.82	2.100	.228
	Male	93	2.55	1.874	.194
Overall score p27q6	Female	85	2.34	2.124	.230
	Male	92	1.91	1.808	.188
Pattern Maintenance p24 q1 or p22 q1	Female	84	5.05	2.454	.268
	Male	90	5.34	2.527	.266
Pattern maintenance p24 q2 or p22 q2	Female	85	4.69	2.673	.290
	Male	90	4.31	2.638	.278
Pattern maintenance p22 q3	Female	44	3.86	3.077	.464
	Male	40	3.20	2.775	.439
Pattern maintenance p22 q4	Female	44	3.50	2.501	.377
	Male	40	3.05	2.353	.372
Score p24 q3	Female	84	3.17	1.387	.151
	Male	88	2.69	1.342	.143
Score p13	Female	84	2.96	.752	.082
	Male	90	2.77	.900	.095
Overall notes sight-reading ability p20	Female	85	1.28	1.342	.146
	Male	92	1.16	1.278	.133
Overall rhythm sight-reading ability p20	Female	85	1.09	1.525	.165
	Male	92	.95	1.354	.141

The group statistics shows the number of male and female participants for each question. Note the numbers are different because some questions were not answered by students and are therefore not valid. Out of the four questions, the only question year 4 and year 8 students had different tasks to complete was task two (pattern maintenance), hence the different numbers of participants.

The mean score of females are higher in all questions except the first question of question two.

In the original NEMP results, it was found year 8 females performed better than males in task two. year 4 females scored higher in task three than males. T-tests were employed to test for significance.

By looking at the independent samples test table, in the column labelled Sig. (2-tailed), females scored significantly higher than males in:

(<.01) in p27's question 2 (0.001), question 3 (0.022), and

(<.05) in p27's question 6 (0.017), and p13 overall score (0.012),

(<.10) in pattern maintenance p22/24 question 3 (0.057) and Overall rhythm sight-reading ability on p20 question.

These results show there are significances among these variables.

Comparing these results with the original NEMP results, in the original NEMP results, girls scored higher than boys did in task 3 (p13), and results in this sample shows there is indeed a significance of 0.012 at 0.05 level. The original result shows year 8

girls scoring higher than year 8 boys in Keyboard Patterns/Rhythm task (task two), in this sample, the results showed a significance if found in year 4 girls scoring higher than year 4 boys in task two question 3, but no significance found between year 8 boys and girls. This t-test was run to see if the 178 sample chosen is representative of the original 2880 sample. From this result, it is not possible to validate whether this is a representative sample of the whole population.

ANOVA was then run to further investigate and evaluate the differences between the mean NEMP scores of male and female.

Female and male scores/results were run, to again compare the sample in this research to the NEMP sample. Still, there was not enough evidence to say with great confidence that it is a representative sample.

Table Six

Independent Samples Test

T-test analysis for mean of students' NEMP scores		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Overall score p27q1	Equal variances assumed	1.429	.233	.972	176	.333	.329	.339	-.340	.998
	Equal variances not assumed			.975	175.971	.331	.329	.338	-.337	.996
Overall score p27q2	Equal variances assumed	10.644	.001	1.371	176	.172	.641	.467	-.281	1.562
	Equal variances not assumed			1.364	169.020	.174	.641	.470	-.286	1.567
Overall score p27q3	Equal variances assumed	5.379	.022	.638	176	.524	.264	.413	-.552	1.079
	Equal variances not assumed			.634	168.556	.527	.264	.415	-.556	1.083
Overall score p27q4	Equal variances assumed	1.251	.265	1.598	176	.112	.787	.492	-.185	1.758
	Equal variances not assumed			1.600	175.195	.111	.787	.492	-.184	1.757
Overall score p27q5	Equal variances assumed	1.239	.267	.924	176	.357	.275	.298	-.313	.863
	Equal variances not assumed			.919	169.024	.359	.275	.299	-.316	.866
Overall score p27q6	Equal variances assumed	5.789	.017	1.447	175	.150	.428	.296	-.156	1.012
	Equal variances not assumed			1.438	165.571	.152	.428	.298	-.160	1.016
Pattern Maintenance p24 q1 or p22 q1	Equal variances assumed	.017	.896	-.785	172	.433	-.297	.378	-1.043	.449
	Equal variances not assumed			-.786	171.724	.433	-.297	.378	-1.042	.449
Pattern maintenance p24 q2 or p22 q2	Equal variances assumed	.106	.745	.954	173	.341	.383	.402	-.410	1.176
	Equal variances not assumed			.953	172.142	.342	.383	.402	-.410	1.176
Pattern maintenance p22 q3	Equal variances assumed	3.736	.057	1.034	82	.304	.664	.642	-.613	1.940
	Equal variances not assumed			1.039	81.996	.302	.664	.639	-.607	1.934
Pattern maintenance p22 q4	Equal variances assumed	.923	.340	.847	82	.399	.450	.531	-.607	1.507
	Equal variances not assumed			.850	81.898	.398	.450	.530	-.604	1.504
Score p24 q3	Equal variances assumed	.614	.435	2.275	170	.024	.473	.208	.063	.884
	Equal variances not assumed			2.274	168.929	.024	.473	.208	.062	.885
Score p13	Equal variances assumed	6.458	.012	1.566	172	.119	.198	.126	-.051	.447
	Equal variances not assumed			1.576	169.949	.117	.198	.125	-.050	.445
Overall notes sight-reading ability p20	Equal variances assumed	1.075	.301	.606	175	.545	.119	.197	-.269	.508
	Equal variances not assumed			.605	172.163	.546	.119	.197	-.270	.509
Overall rhythm sight-reading ability p20	Equal variances assumed	3.320	.070	.686	175	.494	.148	.216	-.279	.576
	Equal variances not assumed			.683	168.438	.496	.148	.217	-.281	.578

Finally, Chi Square test was run to test whether there is any significance in the difference between male and female attitudes or NEMP scores.

Table Seven

Chi-square analysis for students' interest in doing music at school by gender

Gender * How much do you like doing music at school

Crosstab

			How much do you like doing music at school					Total
			not at all	okay	reasonably	very much	Invalid	
Gender	Female	Count	2	19	26	33	1	81
		Expected Count	1.4	20.3	29.9	28.9	0.5	81.0
	Male	Count	1	23	36	27	0	87
		Expected Count	1.6	21.8	32.1	31.1	0.5	87.0
Total		Count	3	42	62	60	1	168
		Expected Count	3.0	42.0	62.0	60.0	1.0	168.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3.718[a]	4	.446
Likelihood Ratio	4.114	4	.391
Linear-by-Linear Association	1.465	1	.226
N of Valid Cases	168		

a. 4 cells (40.0%) have expected count less than 5. The minimum expected count is .48.

There is no significant difference between males and females in terms of enjoyment of music at school. In the original NEMP results showed girls were more positive than boys in overall attitudes to music. Again, the two respective results are different.

Table Eight

Chi-square analysis for students' Task One Question 2 NEMP scores by gender

Gender * Overall score p27q2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	21.533[a]	7	.003
Likelihood Ratio	25.863	7	.001
Linear-by-Linear Association	2.469	1	.116
N of Valid Cases	178		

a. 8 cells (50.0%) have expected count less than 5. The minimum expected count is .48.

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Overall score p27q1	Female	85	6.95	2.165	.235
	Male	93	6.62	2.340	.243
Overall score p27q2	Female	85	4.45	3.293	.357
	Male	93	3.81	2.939	.305
Overall score p27q3	Female	85	5.19	2.922	.317
	Male	93	4.92	2.589	.268
Overall score p27q4	Female	85	4.89	3.240	.351
	Male	93	4.11	3.315	.344
Overall score p27q5	Female	85	2.82	2.100	.228
	Male	93	2.55	1.874	.194
Overall score p27q6	Female	85	2.34	2.124	.230
	Male	92	1.91	1.808	.188
Pattern maintenance p24 q2 or p22 q2	Female	85	4.69	2.673	.290
	Male	90	4.31	2.638	.278
Pattern Maintenance p24 q1 or p22 q1	Female	84	5.05	2.454	.268
	Male	90	5.34	2.527	.266
Pattern maintenance p22 q3	Female	44	3.86	3.077	.464
	Male	40	3.20	2.775	.439
Pattern maintenance p22 q4	Female	44	3.50	2.501	.377
	Male	40	3.05	2.353	.372
Score p24 q3	Female	84	3.17	1.387	.151
	Male	88	2.69	1.342	.143
Score p13	Female	84	2.96	.752	.082
	Male	90	2.77	.900	.095
Overall notes sight-reading ability p20	Female	85	1.28	1.342	.146
	Male	92	1.16	1.278	.133
Overall rhythm sight-reading ability p20	Female	85	1.09	1.525	.165
	Male	92	.95	1.354	.141

Only p27q2 was found with chi-square value 21.533, significant at the 0.003 level. This is the only question that the null hypothesis is rejected with a high degree of confidence. Thus, other than p27q2, there is no significant difference between the numbers of female students scoring higher than male students based on the sample used in this research.

4.2 Results of correlation analysis

The Pearson correlation coefficient makes several assumptions:

1. The two variables are measured using interval- or ratio-scaled measures.
2. The two variables have a linear relationship.
3. Variables analysed are from a bivariate normally distributed population, in other words, the population is such that all the observations with a given value of one variable, have values of the second variable that are normally distributed.[106]

The Pearson correlation method was used to measure the strength of the assumed linear relationship between students' attitudes/experiences and their NEMP selected scores.

Significant correlations found in each attitude related questions:

Table Nine

Correlations between students' attitude and their NEMP scores by year

Attitude Questions	**Significance of correlation found**
How much do you like doing music at school?	Year 4: p22q5 (.230) weak relationship at 0.05 level.
How much do you like singing at school?	None.
How much do you like playing music at school?	None.

[106] Ibid. 568-569.

Attitude Questions	**Significance of correlation found**
How much do you like listening to music at school?	Year 8: p24q3 (.258) weak relationship at 0.05 level.
How much do you like dancing/moving to music at school?	Year 8: p27q3 (.238) weak relationship at 0.05 level.
How much do you like making up music at school?	Year 8: p24q2 (-.219) weak negative relationship at 0.05 level.
How much do you like singing outside school?	None.
How much do you like playing instruments outside school?	Year 8: p27q2 (.245) weak relationship at 0.05 level. Year 8: p27q3 (.261) weak relationship at 0.05 level. Year 8: p24q1 (.221) weak relationship at 0.05 level.
How much do you like listening to music outside school?	Year 8: p20 rhythm sight-reading ability (-.249) weak negative relationship at 0.05 level.
How much do you like dancing/moving to music outside school?	Year 4: p27q6 (.232) weak relationship at 0.05 level. Year 8: p27q1 (.224) weak relationship at 0.05 level. Year 8: p27q3 (.228) weak relationship at 0.05 level.
How much do you like making up music outside school?	Year 4: p20 rhythm sight reading ability (.256) weak relationship at 0.05 level.

There were only weak relationships present.

When comparing correlation between each attitude variables, with year 4 children, there were strong relationships between like doing music at school, dancing/moving to music outside school, and like making up music outside school. With year 8 children, there were strong relationships between playing instruments outside school,

listening, dancing/moving to music at and outside school, making up music outside school.

Most of the correlation results were weak. It is possible that the sample used in this research was too small, but it could also mean there was no significant correlation between attitude and the NEMP tasks analysed.

Results were then compared by combining Year 4 and 8 together to create a greater sample size:

Table Ten

Correlations between combining both year students' attitude and their NEMP scores

Attitude Questions	**Significance of correlation found**
How much do you like doing music at school?	Year 4: p22q1 / Year 8: p24q1 (-.174) weak negative relationship at 0.05 level.
How much do you like singing at school?	Year 4/8: p27q3 (-.198) weak negative relationship at 0.01 level. Year 4/8: p27q5 (-.175) weak negative relationship at 0.05 level. Year 4/8: p20 notes sight-reading ability (-.180) weak negative relationship at 0.05 level.
How much do you like playing music at school?	None.
How much do you like listening to music at school?	Year 4: p22q5 / Year 8: p24q3 (.217) weak relationship at 0.01 level.
How much do you like dancing/moving to music at school?	None.
How much do you like making up music at school?	Year 4: p22q2 / Year 8: p24q2 (-.165) weak negative relationship at 0.05 level.
How much do you like singing outside school?	None.

Attitude Questions	**Significance of correlation found**
How much do you like playing instrument outside school?	None.
How much do you like listening to music outside school?	Year 4/8: p27q3 (.157) weak relationship at 0.05 level. Year 4/8: p27q4 (.175) weak relationship at 0.05 level. Year 4/8: p27q6 (.151) weak relationship at 0.05 level. Year 4: p22q1 / Year 8: p24q1 (.239) weak relationship at 0.01 level.
How much do you like dancing/moving to music outside school?	None.
How much do you like making up music outside school?	None.

Correlation between attitude and score was even weaker when combining the two age groups' results together. A positive weak relationship of 0.217 was found between like listening to music at school and score of p27 q3, and 0.239 was found between like listening to music outside school and score of p24 q1/p22 q1 at 0.01 level.

Results relating to experiences in music looking at Year 4 and 8 separately:

Table Eleven

Correlations between students' experiences and their NEMP scores by year

Experiences Questions	**Significance of correlation found**
How often do you sing at school?	None.
How often do you play instruments at school?	None.

Experiences Questions	**Significance of correlation found**
How often do you listen to music at school?	Year 8: p27q5 (.219) weak relationship at 0.05 level. Year 4: p22q3 (-.284) weak negative relationship at 0.05 level. Year 4: p22q4 (-.337) weak negative relationship at 0.01 level.
How often do you dance/move to music at school?	None.
How often do you make up music at school?	None.
How often do you sing outside school?	Year 4: p27q3 (-.314) weak negative relationship at 0.01 level. Year 4: p27q6 (.272) weak relationship at 0.05 level. Year 8: p20 notes sight-reading ability (-.257) weak negative relationship at 0.05 level. Year 8: p20 rhythm sight-reading ability (-.289) weak negative relationship at 0.01 level.
How often do you play instruments outside school?	Year 4: p27q6 (.281) weak relationship at 0.01 level. Year 4: p20 notes sight-reading ability (.251) weak relationship at 0.05 level.
How often do you listen to music outside school?	Year 4: p20 notes sight-reading ability (-.221) weak negative relationship at 0.05 level. Year 8: p20 rhythm sight-reading ability (-.249) weak negative relationship at 0.05 level.
How often do you dance/move to music outside school?	Year 4: p27q6 (.341) weak relationship at 0.01 level. Year 8: p13 (-.219) weak negative relationship at 0.05 level.

Experiences Questions	**Significance of correlation found**
How often do you make up music outside school?	Year 4: p27q6 (.249) weak relationship at 0.05 level. Year 8: p24q2 (-.253) weak negative relationship at 0.05 level.
Able to identify between high/low sounds?	Year 4: p27q1 (.303) weak relationship at 0.01 level. Year 4: p27q2 (.215) weak relationship at 0.05 level. Year 4: p27q3 (.220) weak relationship at 0.05 level. Year 4: p22q2 (.346) weak relationship at 0.01 level. Year 4: p22q3 (.445) moderate relationship at 0.01 level. Year 4: p20 notes sight-reading ability (.211) weak relationship at 0.05 level. Year 8: p20 notes sight-reading ability (.341) weak relationship at 0.01 level. Year 8: p20 rhythm sight-reading ability (.366) weak relationship at 0.01 level.
Play piano at home/school?	Year 8: p27q2 (.261) weak relationship at 0.05 level. Year 8: p27q2 (.305) weak relationship at 0.01 level. Year 4: p20 notes sight-reading ability (.283) weak relationship at 0.01 level. Year 8: p20 notes sight-reading ability (.302) weak relationship at 0.01 level. Year 8: p20 rhythm sight-reading ability (.314) weak relationship at 0.01 level.
Play other instruments?	Year 4: p27q3 (.221) weak relationship at 0.05 level. Year 4: p20 rhythm sight-reading ability (.301) weak relationship at 0.01 level. Year 8: p13 (.329) weak relationship at 0.01 level.

Experiences Questions	**Significance of correlation found**
Have piano lessons?	Year 4: p27q4 (.286) weak relationship at 0.01 level. Year 4: p27q5 (.412) moderate relationship at 0.01 level. Year 4: p27q6 (.289) weak relationship at 0.01 level. Year 8: p27q5 (.262) weak relationship at 0.05 level. Year 8: p27q6 (.272) weak relationship at 0.05 level. Year 4: p20 rhythm sight-reading ability (.272) weak relationship at 0.01 level. Year 8: p24q2 (.213) weak relationship at 0.05 level. Year 8: p20 notes sight-reading ability (.393) weak relationship at 0.01 level. Year 8: p20 rhythm sight-reading ability (.361) weak relationship at 0.01 level.
Had piano lessons?	Year 4: p27q1 (.214) weak relationship at 0.05 level. Year 4: p27q2 (.253) weak relationship at 0.05 level. Year 4: p27q4 (.306) weak relationship at 0.01 level. Year 4: p27q5 (.432) moderate relationship at 0.01 level. Year 4: p27q6 (.362) weak relationship at 0.01 level. Year 8: p27q1 (.242) weak relationship at 0.05 level. Year 8: p27q2 (.273) weak relationship at 0.05 level. Year 8: p27q3 (.234) weak relationship at 0.05 level. Year 8: p27q5 (.299) weak relationship at 0.01 level. Year 8: p27q6 (.311) weak relationship

Experiences Questions	**Significance of correlation found**
Cont' Had piano lessons?	at 0.01 level. Year 8: p24q2 (.273) weak relationship at 0.05 level. Year 8: p20 notes sight-reading ability (.480) moderate relationship at 0.01 level. Year 8: p20 rhythm sight-reading ability (.422) moderate relationship at 0.01 level.
Learn music or belong to music group outside school?	Year 4: p27q1 (.294) weak relationship at 0.01 level. Year 4: p27q3 (.244) weak relationship at 0.05 level. Year 4: p27q5 (.220) weak relationship at 0.05 level. Year 4: p27q6 (.224) weak relationship at 0.05 level. Year 8: p27q1 (.315) weak relationship at 0.01 level. Year 8: p27q3 (.229) weak relationship at 0.05 level. Year 8: p27q5 (.235) weak relationship at 0.05 level. Year 8: p27q6 (.328) weak relationship at 0.01 level. Year 4: p22q5 (.237) weak relationship at 0.05 level. Year 4: p13 (.244) weak relationship at 0.05 level. Year 4: p20 notes sight-reading ability (.366) weak relationship at 0.01 level. Year 4: p20 rhythm sight-reading ability (.230) weak relationship at 0.05 level.

In the results of experiences questions, there were a greater number of weak relationships present than the attitude questions. In addition, four moderate

relationships were found.

A moderate relationship was found between year 4 students' results in p22q3 and abilities to identify between high/low sounds with 99% confidence. The moderate relationship also exists between year 4 students used to have regular piano lessons and p27 question 5 with 99% confidence. It is not surprising to find that a moderate relationship was found present between used to have regular piano lessons and year 8 students' notes and rhythm sight-reading abilities of p20 questions with 99% confidence, because it is reasonable to assume students who have had regular piano lessons should perform better in tasks requiring skills learned in a piano lesson.

Note that students currently having piano lessons would also count as 'having had/used to have' piano lessons.

Again combining the two years together for the experiences questions:

Table Twelve

Correlations between combining both year students' experiences and their NEMP scores

Experiences Questions	**Significance of correlation found**
How often do you sing at school?	None.
How often do you play instruments at school?	None.
How often do you listen to music at school?	None.
How often do you dance/move to music at school?	None.
How often do you make up music at school?	None.

Experiences Questions	**Significance of correlation found**
How often do you sing outside school?	Year 4/8: p27q3 (-.227) weak negative relationship at 0.01 level. Year 4/8: p27q4 (-.171) weak negative relationship at 0.05 level. Year 4: p22q1 / Year 8: p24q1 (-.181) weak negative relationship at 0.05 level. Year 4/8: p20 notes sight-reading ability (-.166) weak negative relationship at 0.05 level. Year 4/8: p20 rhythm sight-reading ability (-.178) weak negative relationship at 0.05 level.
How often do you play instruments outside school?	Year 4/8: p20 notes sight-reading ability (.198) weak relationship at 0.01 level.
How often do you listen to music outside school?	None
How often do you dance/move to music outside school?	Year 4/8: p13 (-.173) weak negative relationship at 0.05 level. Year 4/8: p20 notes sight-reading ability (-.157) weak negative relationship at 0.05 level.
How often do you make up music outside school?	Year 4: p22q2 / Year 8: p24q2 (-.153) weak negative relationship at 0.05 level.
Able to identify between high/low sounds?	Year 4/8: p27q1 (.310) weak relationship at 0.01 level. Year 4/8: p27q2 (.199) weak relationship at 0.01 level. Year 4/8: p27q3 (.301) weak relationship at 0.01 level. Year 4/8: p27q4 (.297) weak relationship at 0.01 level. Year 4/8: p27q5 (.276) weak relationship at 0.01 level. Year 4/8: p27q6 (.219) weak relationship at 0.01 level. Year 4: p22q1 / Year 8: p24q1 (.255) weak relationship at 0.01 level.

Experiences Questions	**Significance of correlation found**
Cont' Able to identify between high/low sounds?	Year 4: p22q2 / Year 8: p24q2 (.332) weak relationship at 0.01 level. Year 4: p22q5 / Year 8: p24q3 (.289) weak relationship at 0.01 level. Year 4/8: p13 (.235) weak relationship at 0.01 level. Year 4/8: p20 notes sight-reading ability (.218) weak relationship at 0.01 level. Year 4/8: p20 rhythm sight-reading ability (.184) weak relationship at 0.05 level.
Play piano at home/school?	Year 4/8: p27q2 (.181) weak relationship at 0.05 level. Year 4/8: p27q3 (.159) weak relationship at 0.05 level. Year 4/8: p27q5 (.196) weak relationship at 0.01 level. Year 4/8: p27q6 (.245) weak relationship at 0.01 level. Year 4/8: p20 notes sight-reading ability (.315) weak relationship at 0.01 level. Year 4/8: p20 rhythm sight-reading ability (.284) weak relationship at 0.01 level.
Play other instruments?	Year 4/8: p27q2 (.168) weak relationship at 0.05 level. Year 4/8: p20 notes sight-reading ability (.218) weak relationship at 0.01 level. Year 4/8: p20 rhythm sight-reading ability (.257) weak relationship at 0.01 level.
Have piano lessons?	Year 4/8: p27q1 (.179) weak relationship at 0.05 level. Year 4/8: p27q2 (.158) weak relationship at 0.05 level. Year 4/8: p27q4 (.228) weak relationship at 0.01 level.

Experiences Questions	**Significance of correlation found**
Cont' Have piano lessons?	Year 4/8: p27q5 (.320) weak relationship at 0.01 level. Year 4/8: p27q6 (.270) weak relationship at 0.01 level. Year 4: p22q2 / Year 8: p24q2 (.165) weak relationship at 0.05 level. Year 4/8: p20 notes sight-reading ability (.319) weak relationship at 0.01 level. Year 4/8: p20 rhythm sight-reading ability (.330) weak relationship at 0.01 level.
Had piano lessons?	Year 4/8: p27q1 (.234) weak relationship at 0.01 level. Year 4/8: p27q2 (.276) weak relationship at 0.01 level. Year 4/8: p27q3 (.205) weak relationship at 0.01 level. Year 4/8: p27q4 (.265) weak relationship at 0.01 level. Year 4/8: p27q5 (.355) weak relationship at 0.01 level. Year 4/8: p27q6 (.314) weak relationship at 0.01 level. Year 4: p22q2 / Year 8: p24q2 (.182) weak relationship at 0.05 level. Year 4/8: p20 notes sight-reading ability (.365) weak relationship at 0.01 level. Year 4/8: p20 rhythm sight-reading ability (.344) weak relationship at 0.01 level.
Learn music or belong to music group outside school?	Year 4/8: p27q1 (.207) weak relationship at 0.01 level. Year 4/8: p27q2 (.244) weak relationship at 0.01 level. Year 4/8: p27q3 (.197) weak relationship at 0.05 level.

Experiences Questions	Significance of correlation found
Cont' Learn music or belong to music group outside school?	Year 4/8: p27q5 (.185) weak relationship at 0.05 level. Year 4/8: p27q6 (.228) weak relationship at 0.01 level. Year 4: p22q2 / Year 8: p24q2 (.165) weak relationship at 0.05 level. Year 4: p22q5 / Year 8: p24q3 (.171) weak relationship at 0.05 level Year 4/8: p13 (.179) weak relationship at 0.05 level. Year 4/8: p20 notes sight-reading ability (.242) weak relationship at 0.01 level. Year 4/8: p20 rhythm sight-reading ability (.179) weak relationship at 0.05 level.

There was no significant moderate relationship present when combining both years, the reason for that could very possibly be due to the different trends for different years, therefore the original intention of creating a larger sample size did not work simply by combining the two years on the same or similar level questions.

Note that results for questions p22q3 and q4 were not used when combining results because while year 4 student results were available, there were no corresponding year 8 results for comparison. Year 4 students had two more questions to complete.

4.3 Results of chi-square analysis

Chi-square analysis allows us to test for significance between the frequency distributions of two or more groups.

There are some significances found in comparing students' experiences and NEMP results listed as follows:.

1. The following table demonstrates the chi-square analysis for year 4 students recognizing high/low sounds and their NEMP scores for p22q3. It shows the chi-square value of 17.746 is significant at the 0.007 level. Since this level of significance is less than .05, the null hypothesis is rejected with a high degree of confidence. Thus there is a significant difference between year 4 students being able to identify high/low sounds and their pattern maintenance score on p22q3.

Table Thirteen

Chi-square analysis for year 4 students recognising high/low sounds and their NEMP scores for Task Two Question 3

Recognise High/Low sounds * Pattern maintainance p22 q3 * Year

Crosstab

Count

Year			Pattern maintenance p22 q3							Total
			1	2	3	4	6	7	8	
Year 4	Recognise High/Low sounds	No	18	3	4	1	1	0	0	27
		Yes	19	2	7	2	2	5	19	56
	Total		37	5	11	3	3	5	19	83

Chi-Square Tests

Year		Value	df	Asymp. Sig. (2-sided)
Year 4	Pearson Chi-Square	17.746[a]	6	.007
	Likelihood Ratio	24.658	6	.000
	Linear-by-Linear Association	16.236	1	.000
	N of Valid Cases	83		

a. 9 cells (64.3%) have expected count less than 5. The minimum expected count is .98.

2. Significance was found in year 8 students playing piano at home/school and their overall rhythm sight-reading ability on answering p20 questions. The chi-square value was 12.308, significant at the 0.031 level. Both year students' experience in having piano lessons was significant on their overall notes sight-reading ability on p20 questions.

3. Year 4 students having piano lessons was significant on their overall scores of p27q5.

4. Both year students having had piano lessons was significant on their overall notes sight-reading ability of p20 questions.

5. Between having had piano lessons and overall rhythm sight-reading ability on p20, a significance was found on year 4 students with 90% confidence, and 95% confidence for year 8 students.

6. Both year students having had piano lessons was significant with their overall scores on p27q5 and q6.

7. When both year students' results were combined, significances were found between them recognising high/low sounds and their scores on pattern maintenance p24q2/p22q2, q3, p27q1, q3, q4, and q5. Significances were found between students playing piano at home/school and their overall notes and rhythm sight-reading ability on p20. Significances were found between students' having experience in playing other instruments and their overall rhythm sight-reading ability on p20 questions. Similarly, students having piano lessons and their overall notes and rhythm sight-reading ability on p20 questions, overall scores on p27q5 and q6 all have shown significances. Students who had lessons show results of significant for their overall notes and rhythm sight-reading ability on p20 questions, and overall scores on p27q1, q2, q4, q5 and q6.

4.4 Results of analysis of variance

Analysis of variance (ANOVA) is used to determine the statistical difference between three or more means. Different scores of individual questions were analysed separately to allow easier data presentation.[107]

Independent variable: Scores of individual NEMP tasks.

Dependent variables: Students' attitudes.

ANOVA tests whether the observed differences in the variance of students' attitudes are statistically significant.

Assumptions for using ANOVA:

1. Dependent variable must be either interval or ratio scaled.
2. Independent variable needs to be categorical.

ANOVA was used to compare students with different attitude/experience scores and their scores in different NEMP tasks.

Null hypothesis:

There is no difference between the NEMP scores of students with different attitude/experience scores.

Alternative hypothesis:

There is a difference between the NEMP scores of students with different attitude/experience scores and is not due to random error.

---

[107] Ibid. 543.

Significances were found in the following variables:

Table Fourteen (a)

ANOVA on students' NEMP scores by different attitudes and experiences

Attitude/Experiences	Questions	F-ratio	Probability level
Like doing music at school	Pattern maintenance p24q1 or p22q1	3.028	.031
Time spent on singing at school	Overall score p27q2 Overall score p27q3	2.480 2.544	.063 .058
Time spent on instrument at school	Overall score p27q6 Pattern maintenance p24q1 or p22q1 Overall notes sight-reading ability p20	2.217 2.194 2.196	.088 .091 .090
Time spent on listening to music at school	Pattern maintenance p22q3 Pattern maintenance p22q4	3.646 4.184	.016 .009
Time spent on dancing/moving to music at school	Score p13	2.289	.080
Like singing at school	Overall score p27q3 Pattern maintenance p22q4	3.567 2.166	.015 .099
Like playing instrument at school	Score p24q3	3.221	.024
Like listening to music at school	Score p24q3 Score p13	3.789 3.071	.012 .029

Attitude/Experiences	Questions	F-ratio	Probability level
Like dancing/moving to music at school	Pattern maintenance p24q1 or p22q1	2.214	.088
	Pattern maintenance p22q3	2.670	.053
	Score p24q3	2.350	.074
Like making up music at school	Overall score p27q1	7.445	.007
	Overall score p27q2	10.492	.001
	Overall score p27q3	6.719	.010
	Overall score p27q4	3.416	.066
	Overall score p27q5	5.897	.016
	Overall score p27q6	9.073	.003
	Score p24q3	4.805	.030
	Score p13	5.367	.022
	Overall notes sight-reading ability p20	10.267	.002
	Overall rhythm sight-reading ability p20	5.464	.021
Time spent on singing outside school	Overall score p27q1	3.248	.023
	Overall score p27q3	4.979	.002
	Overall score p27q4	2.720	.046
	Pattern maintenance p24q1 or p22q1	2.249	.085
	Pattern maintenance p24q2 or p22q2	3.184	.025
	Score p13	3.888	.010
	Overall rhythm sight-reading ability p20	2.490	.062
Time spent on playing instrument outside school	Overall notes sight-reading ability p20	3.142	.027
Time spent on dancing/moving to music outside school	Score p13	2.368	.073

Attitude/Experiences	Questions	F-ratio	Probability level
Like singing outside school	Overall notes sight-reading ability p20 Overall rhythm sight-reading ability p20	2.172 2.556	.093 .057
Like playing instrument outside school	Pattern maintenance p24q1 or p22q1 Pattern maintenance p22q3	2.267 2.477	.083 .068
Like listening to music outside school	Overall score p27q4 Pattern maintenance p24q1 or p22q1	2.862 4.949	.060 .008
Like making up music outside school	Overall score p27q2 Pattern maintenance p24q2 or p22q2	3.514 2.434	.017 .067

Like making up music at school had most statistical difference between students' attitude/experience scores and their NEMP scores. The larger *F* ratios imply significant difference between the groups. The larger the *F* ratio, the more likely it is that the null hypothesis will be rejected.

Table Fourteen (b)

ANOVA

		Sum of Squares	df	Square	F	Sig.
Overall score p27q1	Between Groups	38.290	1	38.290	7.445	0.007
	Within Groups	853.704	166	5.143		
	Total	891.994	167			
Overall score p27q2	Between Groups	97.567	1	97.567	10.492	0.001
	Within Groups	1,543.713	166	9.299		
	Total	1,641.280	167			
Overall score p27q3	Between Groups	48.824	1	48.824	6.719	0.010
	Within Groups	1,206.295	166	7.267		
	Total	1,255.119	167			
Overall score p27q4	Between Groups	36.239	1	36.239	3.416	0.066
	Within Groups	1,761.041	166	10.609		
	Total	1,797.280	167			
Overall score p27q5	Between Groups	22.984	1	22.984	5.897	0.016
	Within Groups	647.010	166	3.898		
	Total	669.994	167			
Overall score p27q6	Between Groups	34.884	1	34.884	9.073	0.003
	Within Groups	634.421	165	3.845		
	Total	669.305	166			
Pattern Maintenance p24 q1 or p22 q1	Between Groups	1.996	1	1.996	0.322	0.571
	Within Groups	1,004.199	162	6.199		
	Total	1,006.195	163			
Pattern maintenance p24 q2 or p22 q2	Between Groups	30.934	1	30.934	4.571	0.034
	Within Groups	1,103.211	163	6.768		
	Total	1,134.145	164			
Pattern maintenance p22 q3	Between Groups	0.035	1	0.035	0.004	0.951
	Within Groups	698.644	76	9.193		
	Total	698.679	77			
Pattern maintenance p22 q4	Between Groups	2.126	1	2.126	0.338	0.563
	Within Groups	477.669	76	6.285		
	Total	479.795	77			
Score p24 q3	Between Groups	9.202	1	9.202	4.805	0.030
	Within Groups	306.403	160	1.915		
	Total	315.605	161			
Score p13	Between Groups	3.759	1	3.759	5.367	0.022
	Within Groups	114.144	163	0.700		
	Total	117.903	164			
Overall notes sight-reading ability p20	Between Groups	16.949	1	16.949	10.267	0.002
	Within Groups	272.380	165	1.651		
	Total	289.329	166			
Overall rhythm sight-reading ability p20	Between Groups	11.113	1	11.113	5.464	0.021
	Within Groups	335.593	165	2.034		
	Total	346.707	166			

4.5 Results of regression analysis

Table Fifteen

Regression analysis for Task One Question One using students' attitudes as predictors

Model Summary(d)

Model	R	R Square	Adjusted R Square	Std. Error of the
1	.270(a)	0.073	0.000	2.334
2	.392(b)	0.154	0.018	2.314
3	.551(c)	0.303	0.153	2.149

a. Predictors: (Constant), Like making up music outside school, Like listening to music outside school, Like singing at school, Like playing instrument at school, Like listening to music at school, Like dancing/moving to music at school, Like doing music

b. Predictors: (Constant), Like making up music outside school, Like listening to music outside school, Like singing at school, Like playing instrument at school, Like listening to music at school, Like dancing/moving to music at school, Like doing music

c. Predictors: (Constant), Like making up music outside school, Like listening to music outside school, Like singing at school, Like playing instrument at school, Like listening to music at school, Like dancing/moving to music at school, Like doing music

d. Dependent Variable: Overall score p27q1

ANOVA(d)

Model		Squares	df	Square	F	Sig.
1	Regression	60.216	11	5.474	1.005	.445(a)
	Residual	768.346	141	5.449		
	Total	828.562	152			
2	Regression	127.265	21	6.060	1.132	.324(b)
	Residual	701.298	131	5.353		
	Total	828.562	152			
3	Regression	251.166	27	9.302	2.014	.005(c)
	Residual	577.396	125	4.619		
	Total	828.562	152			

music outside school, Like singing at school, Like playing instrument at school, Like listening to music at school, Like dancing/moving to music at school, Like doing music

music outside school, Like singing at school, Like playing instrument at school, Like listening to music at school, Like dancing/moving to music at school, Like doing music

music outside school, Like singing at school, Like playing instrument at school, Like listening to music at school, Like dancing/moving to music at school, Like doing music

d. Dependent Variable: Overall score p27q1

Regression analysis helps us understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the average value of the dependent variable when the independent variables are held fixed. Regression analysis is widely used for prediction and forecasting. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships.[108]

Hedden explained in his research that, "the value of these multiple regression studies is that they provide information on the relative importance of certain variables with respect to music achievement. In a sense, these studies serve an exploratory function in that they investigate relationships rather than causal effects."[109]

Multiple regression analysis was used in this situation. The variables were grouped into three categories: attitudes, experiences, and ability and experiences.

Approximately 30.3 percent of the total variation in students' scores is associated with all the variables tested. The above results show significance is only present when considering all of the variables. Attitude variables only represented 7.3 percent of students' p27q1 scores. Many questions showed more variables are required to explain students' scores. The results from the unstandardised regression coefficient have shown there were weak relationships present; significances were found but the coefficients were small.

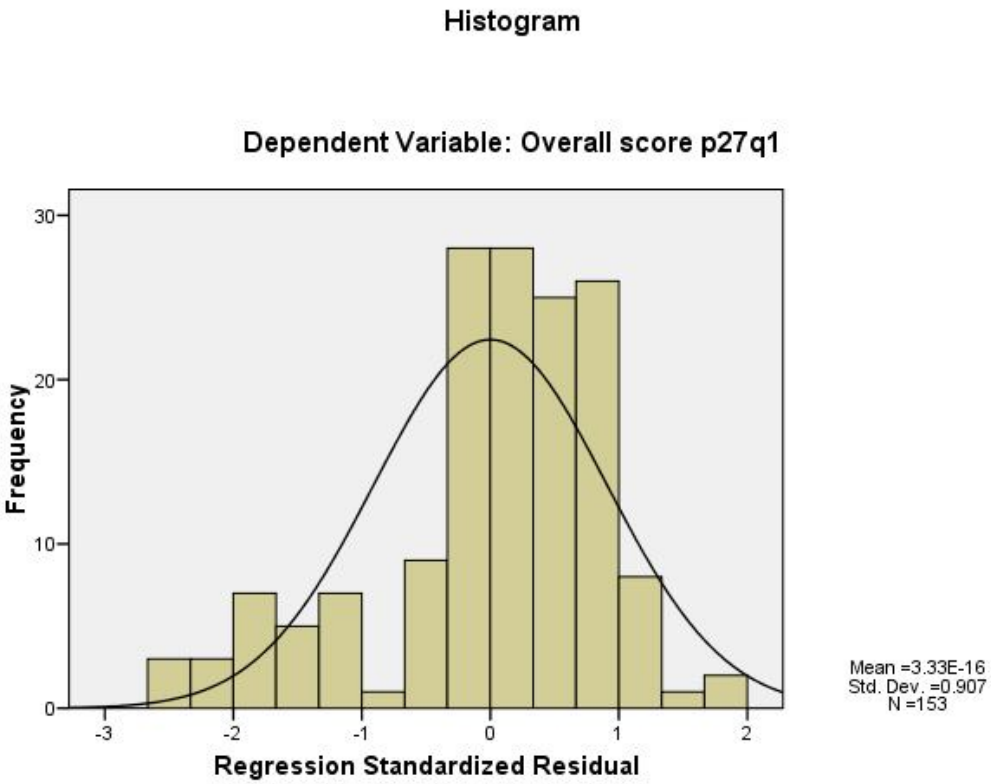
---

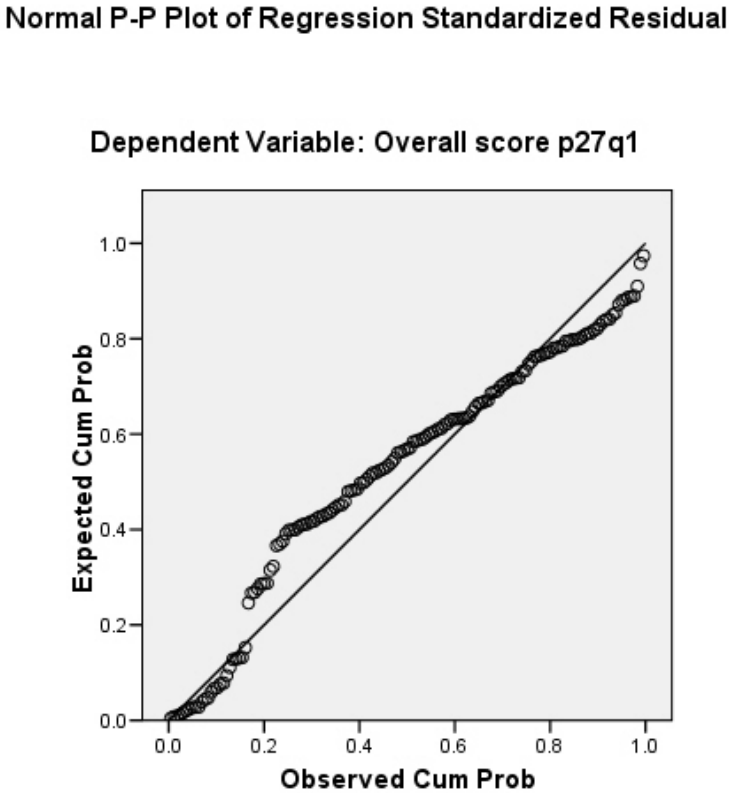
[108] Ibid. 572.

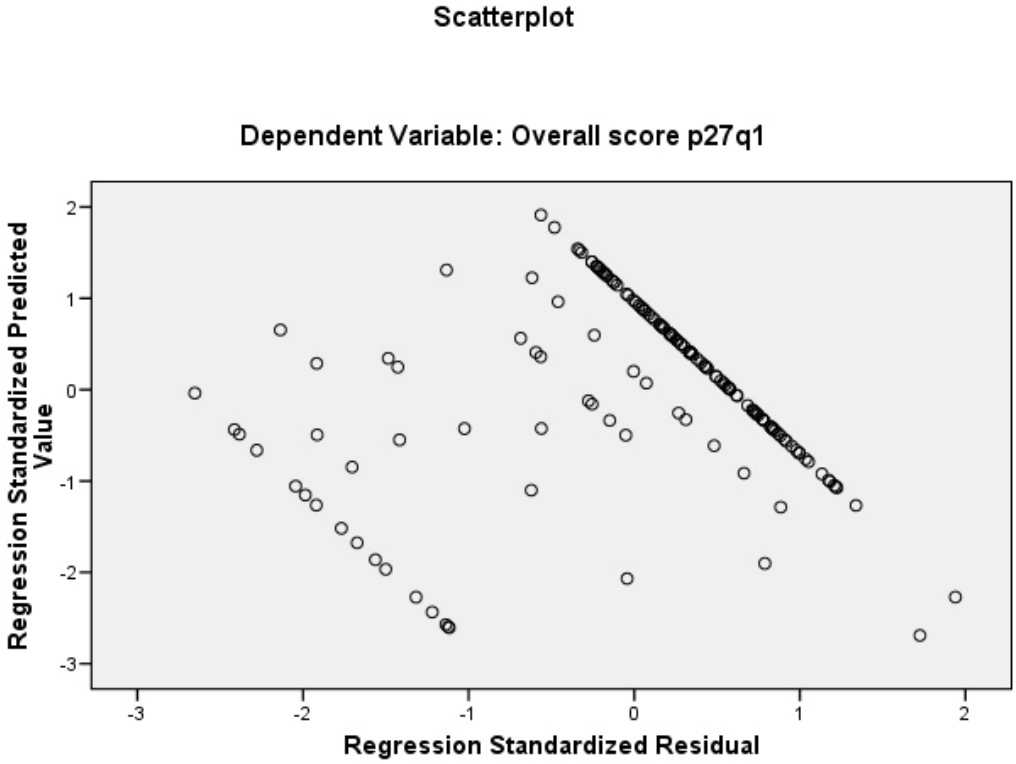
[109] Hedden, S. K. (1982). Prediction of Music Achievement in the Elementary School. *Journal of Research in Music Education, 30*(1), 62.

Figure One

Residual graphs for students' regression residual scores in Task One Question One



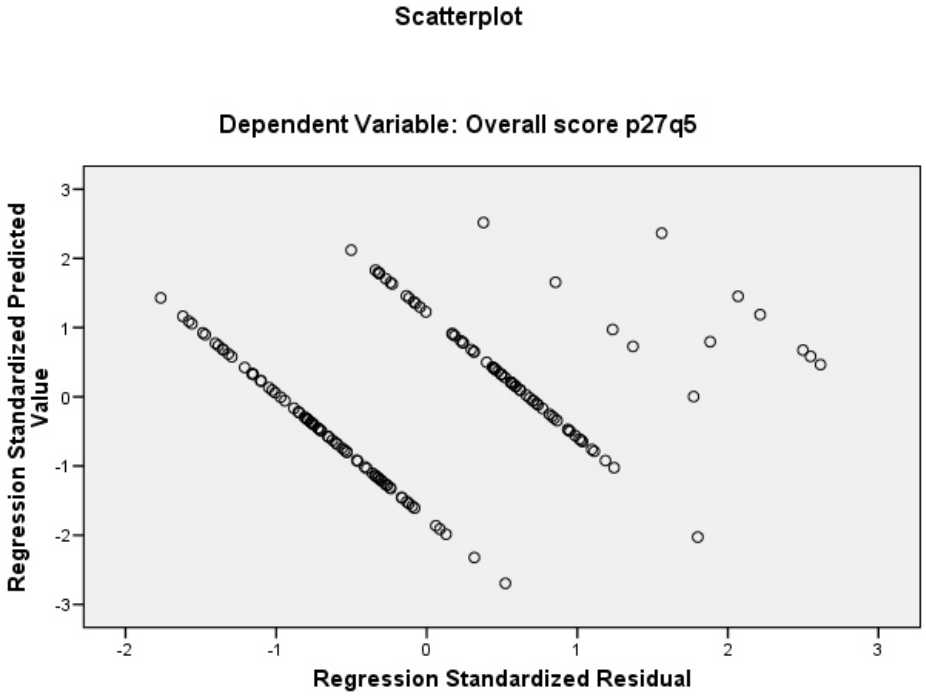




Looking at the above three graphs should help with diagnosing potential problems caused by data observations for students scores in p27q1. The first graph shows some residuals are relatively far away from the curve, which could mean that there might be a problem. Looking at the second graph, if the observed residuals are normally distributed, they will fall directly on the 45-degree line shown on the graph. We can see from our regression model that the residuals are fairly close. Finally, the third graph compares the standardised predicted value of the dependent variable with the standardised residuals from the regression equation. When a model provides a good fit and does not violate any model assumptions, this type of residual plot exhibits no marked pattern or trend. The last graph exhibits no such trend, indicating an adequate fit.

Figure Two

Scatterplot for students' regression residual scores in Task One Question Five



The third scatter plot graph for p27q5 seems to represent an outlier problem. Looking at the following table showing distribution of the scores, 1, 4, 6, 7 and 8, the scores are not spread in an even manner.

Table Sixteen

Descriptive statistic for students' scores in Task One Question Five and the number of students having had lessons by year

Crosstab

Count

Year			Overall score p27q5					Total
			1	4	6	7	8	
Year 4	Had lessons	No	60	11	0		0	71
		Yes	9	8	2		1	20
	Total		69	19	2		1	91
Year 8	Had lessons	No	22	34	1	1	3	61
		Yes	3	16	3	0	4	26
	Total		25	50	4	1	7	87

Table Seventeen

Regression analysis for Task One Question Three using year 8 students' attitudes as predictors

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.449[a]	.202	.088	2.068

a. Predictors: (Constant), Like making up music outside school, Like listening to music outside school, Like dancing/moving to music at school, Like listening to music at school, Like playing instrument at school, How much do you like doing music at school, Like dancing/moving to music outside school, How much do you like singing at school, Like making up music at school, Like playing instrument outside school

b. Year 4 or Year 8 = Year 8

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	75.823	10	7.582	1.772	.082[a]
	Residual	299.461	70	4.278		
	Total	375.284	80			

a. Predictors: (Constant), Like making up music outside school, Like listening to music outside school, Like dancing/moving to music at school, Like listening to music at school, Like playing instrument at school, How much do you like doing music at school, Like dancing/moving to music outside school, How much do you like singing at school, Like making up music at school, Like playing instrument outside school

b. Dependent Variable: Overall score p27q3

c. Year 4 or Year 8 = Year 8

The R square is .202, which means that approximately 20% of the variations in p27q3 year 8 score is associated with the variation in attitudes, significant at 0.1 level. The F-ratio is rather small. The F-ratio is the result of comparing the amount of explained variance to the unexplained variance. The smaller the F-ratio, the less variance in the

dependent variable is associated with the independent variable. The smaller the F ratio, the more likely it is that the null hypothesis will not be rejected.

Table Eighteen

Regression analysis for Task Two Question Two using year 8 students' attitudes as predictors

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.456[a]	.208	.094	2.251

a. Predictors: (Constant), Like making up music outside school, Like listening to music outside school, Like dancing/moving to music at school, Like listening to music at school, Like playing instrument at school, How much do you like doing music at school, Like dancing/moving to music outside school, How much do you like singing at school, Like making up music at school, Like playing instrument outside school

b. Year 4 or Year 8 = Year 8

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	92.014	10	9.201	1.816	.074[a]
	Residual	349.536	69	5.066		
	Total	441.550	79			

a. Predictors: (Constant), Like making up music outside school, Like listening to music outside school, Like dancing/moving to music at school, Like listening to music at school, Like playing instrument at school, How much do you like doing music at school, Like dancing/moving to music outside school, How much do you like singing at school, Like making up music at school, Like playing instrument outside school

b. Dependent Variable: IPattern maintenance p24q2 or p22q2

c. Year 4 or Year 8 = Year 8

21% of the variations in p24 q2 year 8 score is associated with the variation in attitudes with significance at 0.1 level.

21% of the variations in p20 overall rhythm sight-reading ability for year 4 score is associated with the variation in attitudes with significance at 0.1 level.

Table Nineteen

Regression analysis for Task Four overall rhythm sight-reading ability score using year 4 students' attitudes as predictors

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.462[a]	.213	.096	1.006

a. Predictors: (Constant), Like making up music outside school, Like listening to music outside school, Like dancing/moving to music at school, Like playing instrument at school, Like listening to music at school, Like playing instrument outside school, How much do you like singing at school, Like making up music at school, How much do you like doing music at school, Like dancing/moving to music outside school

b. Year 4 or Year 8 = Year 4

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18.371	10	1.837	1.814	.075[a]
	Residual	67.847	67	1.013		
	Total	86.218	77			

a. Predictors: (Constant), Like making up music outside school, Like listening to music outside school, Like dancing/moving to music at school, Like playing instrument at school, Like listening to music at school, Like playing instrument outside school, How much do you like singing at school, Like making up music at school, How much do you like doing music at school, Like dancing/moving to music outside school

b. Dependent Variable: Overall rhythem sight-reading ability p20

c. Year 4 or Year 8 = Year 4

4.6 Reliability, validity and limitations

Limitations of the research design

As previously mentioned in chapter three, the results of this research can only represent the sampled population, and do not represent the overall population of year 4 and 8 New Zealand students.

Some students have not completed the questions or surveys, their results are still included in the analysis to take into account the limitations of the NEMP data.

The researcher analysed the video tapes in this research based partly on NEMP marking criterion, and partly on her own criteria, which were specifically designed for the statistical analyses purpose of this research. Having the same person giving marks for each student should ensure consistency in student scores.

In the survey section, students were asked to answer a questionnaire related to their attitude and experience with music in and out of school with the use of the Likert scale. The limitation lies in the differing interpretation of the different description of the scales.

Due to original NEMP data not being collected by the researcher, inconsistency in teacher administrators was uncontrollable. Although teacher administrators were given instructions to ensure consistency between each child, there were still

differences between each administrator's approaches, which might have affected some of the results unintentionally. Some of the instructions probably require more attention to improve the consistency problem.

The students who had encountered more encouraging teacher administrators generally attempted more questions, which helped the students to understand instructions better and therefore increased chances in obtaining answers that are more accurate. These behaviours have influenced some of the results.

The incompletion and misinterpretation of some of the surveys is another point that should be noted. Although the teacher administrators were asked to check the completion of all the survey forms, some were very incomplete and some were filled in a way which showed some children did not understand the survey form. This shows that a percentage of the survey may not have been conducted with enough care.

Limitations

The results of this research are limited to the defined population of this study, that is, children in year 4 and year 8 of New Zealand schools.

Learning is a difficult element to measure. People are born with different physical abilities. They might understand something but are unable to perform it. This aspect is briefly discussed along with ability in chapter two. For the purpose of this research, students were given credit for demonstrating required skills.

Although different assessors were involved in the original NEMP marking, consistency was achieved through common instructions and marking criteria.

Another potential variable is in the administration of the NEMP tasks. This was minimised by employing specially trained teacher administrators. However, for the purpose of this study, the researcher examined the children's performances of the tasks directly, and employed her own criteria that were apposite to the demands of the study.

Validity types and threats to validity:

Threats to internal validity

History

History is not controlled. Students did the assessment in different environment depending on their schools, and some were noisy. Some of the tests were interrupted by school breaks or end of school time.

Maturation

Not a concern for this research, since most students are in the same year and the same approximate age.

Testing

Students more or less complete the test within a day, therefore testing should not be a validity threat.

Instrumentation

The tapes were all re-analysed by the researcher to ensure consistency in marking, therefore avoid different interpretation of marking criterions.

Selection bias

Random sampling was used in this research to avoid selection bias.

Statistical regression

A possible threat exists in the extreme results. Some students had understanding difficulty, because they were given the choice of not required to answer if they did not know, some gave up more easily.

Mortality

Although 120 schools at year 4 and 120 at year 8 levels were selected randomly, schools were given the option of whether to participate or not in the assessment. In 2000, 296 schools were invited to participate, and 291 agreed. All five schools that declined were in the year 8 sample. Within the student selection, some were replaced due to problems such as limited English language skills, disabilities, or other problems. All of these could have affected the validity of the randomization.

Ambiguity

Ambiguity exists in the inconsistency of teacher administrators' guidelines given to children. It is hard to differentiate whether students' achieved higher because they were given help to understand the question more, or because positive encouraging environment promotes better testing results.

Threats to External Validity

Treatment vs. treatment

Students' were assessed by different administrators, in other words, they could have respond differently if they were assessed by different people. Some might be more encouraging. The administrators were following some of the guidelines differently, probably due to personal interpretations, although this problem was attempted to minimise by group session in Wellington to train these administrators, it was still obvious on the video tapes that there is a validity problem.

Treatment vs. testing

Hawthorne effect – students might act differently knowing they are being assessed.

Treatment vs. selection

Sampled schools were chosen randomly to eliminate the possibility of choosing schools within the same demographics. Still, it is possible students within different graphics would have different achievement and interest behaviours.

Treatment vs. setting

Students were tested in different environments. It is possible students would achieve better when they were given a quieter environment.

Treatment vs. history

The assessments were all conducted during school days. The only possible threat of interactions of treatment with history would be everyone else in the assessed student's class doing other activities which might distract the assessed student.

Construct Validity

Mono-operation and monomethod bias

More than one method of measuring outcomes of the dependent variable was used to avoid this problem. More than one measuring instruments were used. Students were tested by four approaches, namely one-to-one interview, stations, team, and group and independent. This was already discussed in chapter one.

Hypothesis-guessing

The purpose in this research was set out to find if there's a relationship, and if so, what type of relationship, rather than assume one definitely affects the other.

Evaluation apprehension

Anonymity of students were ensured. Different groups of people were selected to be teacher administrators and final statistic analysts.

Demand characteristics

Some administrator encouraged students to guess more, rather than following the original NEMP instruction to ask the student to discontinue when they feel they can not do the assessed tasks. There is a problem in determining how much help the administrators are allowed to give to the assessed students. Because of this, some answers might not be representative of what the assessed student could actually do, as in their actual abilities.

Diffusion of treatment

Students, who were assessed first, might discuss their experiences with other students.

Reliability

Consistency were ensured in this research by using researcher's own set of marking criterions. Future research should become more reliable if more emphasis can be placed on the training of teacher administrators; ensure same extent of help is given to assessed students.

Chapter V – Discussion and conclusion

5.1 Discussion of results and implications for music education

This research was conducted initially with the aim to see if there is a relationship between student attitude and achievement. It is important to bear in mind that correlation is not causation. Correlation cannot prove a causal relationship, but can suggest possible influence from one to the other.

As measured by correlation analyses, based on the sample used in this research, results seem to suggest that attitudes towards music do not have a strong influence on students' musical abilities. In both situations when analysing either Year 4 or Year 8 separately or together, the results have shown that only few weak or no significant correlations are in place..

It seems logical to assume:

- task one - *Coloured Keys* might be relevant to attitude in relation to listening to music or playing instruments;
- task two - *Keyboard Patterns/Rhythms* might have correlation with playing instruments, listening to music, or making up music;
- task three - *Fun Day* might be influenced by attitude on singing, playing instruments and making up music; and
- task four - *Play It* might be affected by attitudes on playing instruments.

The results of the time spent on music relating aspects in/out school was subsequently assessed due to the lack of results in the comparison of attitudes and scores. There

was no strong correlation found in this subsequent assessment also.

When correlation coefficient is weak, two possibilities must be considered:

1. there is no consistent, systematic relationship between the two variables in the population; or
2. the association exists, but it is not linear, and other types of relationships must be investigated further.

In Pearson's correlation analysis, although a greater number of relationships were found between musical experiences and students' scores, the relationships were mostly weak with the exceptions of four being moderate. Again, this is similar finding to similar researches on attitude.

In Sylvia Estes Cary's 1981 Ph.D. thesis[110], the population for her study is defined as fifth grade students in public schools in Oregon that have low minority level and predominantly middle class to upper-middle class socio-economic level. All of the fifth grade students from the experimental school and from the control school were given the pre-tests and post-tests of the *Music Achievement Tests* by Richard Colwell. The measures were the independent variables for the individualised music instruction group and the traditional music instruction group – music achievement (pre and post tests), music performance (pre and post tests), music aptitude, music attitude, and academic reading.

---

[110] Cary, S. E. (1981). *Individualised Music Instruction – Traditional Music Instruction: Relationships of Music Achievement, Music Performance, Music attitude, Music Aptitude, and Reading in Classes of Fifth Grade Students.* Unpublished doctoral dissertation. Oregon University, Michigan.

She found that attitude has the least influence on achievement in comparison with other variables. Overall, attitude had fewer significant correlations than any other variable. Also, in the multiple regression analysis attitude had a minor influence. Cary was surprised by these findings because of teachers' tendency to believe that attitude is an important correlation with achievement. The experimental group, whose achievements and performances were higher, had no significant correlation with attitude. The control group with lower achievement and performances had some significant correlations with attitude. Cary found that the analysis of variance results shows that, the experimental group had a significantly better attitude and made the relationship between attitude and achievement 'thought provoking'. It is interesting that in this research students' experiences in music have shown more significance on their scores than students' attitudes toward music. The sample in this research is not large enough to divide students into different level of achievement, otherwise it would be interesting to see if it matches what Cary has found.

Although this research did not find significant relationship between attitude and achievement, it does not mean attitude should be overlooked in schools, because the result of this research only represent the small sampled 178 students.

Similar to Cary, Steven K. Hedden's research used Cowell's (1969) *Music Achievement Tests,* as well as *Iowa Tests of Basic Skills, Attitude toward Music Scale,* and Svengalis' *Self-concept in Music Scale* in his research. His sample population was fifth and sixth graders in two small Midwestern towns. Instead of using correlation analysis, Hedden ran regression analysis to see how much each variable weighted as a predictor. He conducted the analysis on two schools. At school one, a value of R=.58 was obtained by having academic achievement and self-concept in music as

predictors. In school two, a value of R=.78 was obtained when academic achievement and attitude toward music were used as the two significant predictors. In his research, he found self-concept in music, and attitude might be a predictor for students' academic achievement in music. This is the opposite result of this research and Cary's finding. Still, it is likely the samples used in these researches are simply not representative enough. The regression analysis of this research showed attitude variables only acted as a small predictor for students' scores.

ANOVA has shown that there is a difference between scores for people who currently have piano/keyboard lessons. This is an expected result, because the four given questions in the video tapes were all in some ways keyboard related. Although the result for difference in individual experiences were not great, the analysis showed when all experiences were taken into account, children with greater exposure and greater interest to music do seem to give better musical performances.

From the results of this research, the experiences variables have shown more significant results than attitude variables. It would seem attitude might not be as important as predicted. Nevertheless, attitude may still play an important role in the education curriculum. Many educationists observed attitude as an important element in teaching situations; Val Drew said, "When learning is a pleasurable experience, a window of opportunities await…."[111] Clearly, further study in this area is warranted to be certain of how much influence attitude has on students.

Although low-stake testing is in general more time consuming for everyone

[111] Drew, V. (1965). *Learning to Sing can be Fun.* Christchurch: Dunford.

concerned, the fact that there are countries, which are willing to spend the extra effort in adopting a low-stake testing method affirm its importance. NEMP assessment has only been in practice for 17 years, many areas are still in need of improvement to ensure its reliability and consistency while conducting the assessments. Out of all the available national formal assessments, the way in which NEMP surveys measure attitude is unique. There are still many remaining issues when analysing human behaviour and therefore affects the validity of measuring attitude. More in depth research is clearly required in order to extend existing knowledge on potential attributes to students' achievement to provide a more useful assessment strategy for one's education system.

5.2 Future recommendations

This study should be of value to the development of the NEMP assessments, as well as to educationists. It is recommended that this research be replicated with a same sample size but different population. There were schools within different socio-economic classes taken, it is possible student behaviour in different socio-economic classes might have diverse correlation patterns.

Although time and money will be required for a larger sample size, it is recommended for the reliability of this research.

Improvement on the original NEMP data collection is also recommended. As previously mentioned in the limitation of research section, students had varied support in answering the questions. This inconsistency can only be improved by increased guidance and further training of the teacher administrators.

It is recommended further research to include 2004 and 2008 results to see if any trends are present with the link tasks.

It might also be beneficial to look at effects of attitude in classrooms from a psychology perspective across all school subjects. It would be interesting to see whether different physical elements required for different subjects would give different results.

Bibliography

ACCAC (Qualifications, Curriculum and Assessment Authority for Wales) (2004). *Review of the School Curriculum and Assessment Arrangements 5-16.* Cardiff: ACCAC.

Allport, G. W. (1935). *Attitudes. In C. Murchison (Ed.), A handbook of social psychology.* Worchester, Mass.: Clark University Press.

Andrews, C. Brown, R. et al., (2007). Compulsory assessment systems in the INCA countries: Thematic Probe, *INCA Internet Archive*.

Bentley, A. (1966). *Musical Ability in Children and its Measurement.* New York: October House.

Braatvedt, S. P. (1998). *The Role of Singing in the NEMP Tests.* Unpublished master's thesis dissertation, University of Canterbury, Christchurch, New Zealand.

Broquist, O. A. (1961). *A Survey of the Attitude of 2,594 Wisconsin Elementary School Pupils towards their Learning Experience in Music,* Doctoral dissertation, University of Wisconsin, Madison, WI.

Buckton, R. M. (1981). *The Development of Musical Concepts in Young Children: an Investigation.* Unpublished doctoral dissertation. University of Auckland, Auckland, New Zealand.

Buckton, R. M. (1983). Sing a Song of Six-Year-Olds. *New Zealand Council for Educational Research,* Wellington. 10.

Bullock, W. J. (1975). Student Attitude and Musical Response, *Journal of Aesthetic Education, 9*(4). 109-112.

Cary, S. E. (1981). *Individualised Music Instruction – Traditional Music Instruction: Relationships of Music Achievement, Music Performance, Music attitude, Music Aptitude, and Reading in Classes of Fifth Grade Students.* Unpublished doctoral dissertation. Oregon University, Michigan.

Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement, Issues and Practice*, *20*(4). 19-28.

Colwell, R. (1969). *Music Achievement Tests.* Chicago: Follett Educational Corporation.

Colwell, R. (1970). *Interpretive Manual Music Achievement Tests 1 and 2.* Chicago: Follett.

Crickmore, L. (1968). An Approach to the Measurement of Music Appreciation. *Journal of Research in Music Education*, *16*(3). 239-253.

Crooks, T. J. (2002). Educational Assessment in New Zealand Schools, *Assessment in Education, 9*(2).

Department of Education. (1989). *New Zealand Syllabus for Schools of Music Education – Early Childhood to Form Seven.* Wellington, N.Z.: Author.

Drew, V. (1965). *Learning to Sing can be Fun.* Christchurch: Dunford.

Edwards, J. S & Edwards, M. C. (1971). A Scale to Measure Attitudes toward Music. *Journal of Research in Music Education, 19*(2). 228-233.

English, H. B. & English, A. C. (1958). *A comprehensive dictionary of psychological and psychoanalytic terms: A guide to usage.* New York: McKay.

Fishbein, M. (1967). *Attitude and the prediction of behavior. In M. Fishbein (Ed.),* Readings in attitude theory and measurement. New York: Wiley.

Flexer, R.J., Cumbo, K., Borko, H., Mayfield, V & Maion, S.F. (1995). How 'messing about' with performance assessment in mathematics affects what happens in classrooms. *Technical Report, 396, Los Angeles Centre for Research on Evaluation, Standards and Student Testing (CRESST).*

Gilmore, A. (1999). NEMP: Evaluation of teachers' professional development. In M. Kane (Chair), The National Monitoring Project: Validity, professional development, and group assessment processes. Symposium conducted at the annual meeting of the American Educational Research Association, Montreal, Canada.

Gordon, E. E. (1965). *Musical Aptitude Profile.* Boston: Houghton Mifflin.

Gordon, E. E. (1980). *Learning sequences in Music: Skill, Content, and Patterns*. (1984 edi.). Chicago: IL: GIA.

Gordon, E. E. (1987). *The Nature, Description, Measurement, and Evaluation of Music Aptitudes*. Chicago: IL: GIA.

Hall, K. & Harding, A. (2002). Level descriptions and teacher assessment in England: towards a community of assessment practice. *Educational Research*. 1-15.

Hall, K., Webber, B., Varley, S, Young, V & Dorman, P. (1997). A study of teachers' assessment at Key Stage 1. *Cambridge Journal of Education, 27*. 107-122.

Hair et al. (2003) *Marketing Research – Within a changing Information Environment* (2nd ed.). New York, NY: McGraw-Hill.

Harlen, W. (2007). The quality of learning: assessment alternative for primary education. *Primary Review Research Survey 3/4,* Cambridge: University of Cambridge Faculty of Education.

Hedden, S. K. (1982). Prediction of Music Achievement in the Elementary School. *Journal of Research in Music Education, 30*(1). 61-68.

Jones. L. V. (2001). Assessing achievement versus high-stakes testing: a crucial contrast. *Educational Assessment*, *7*(1). 21-28.

King, A. (1983). Agency, Achievement, and Self-Concept of Young Adolescent Art Students. *Studies in Art Education*, *24*(3). 187-194.

Kirchhubel, J. (2002). *Adolescent Music Development and the Influence of Pre-Tertiary Specialised Music Training.* Unpublished doctoral dissertation. Griffith University, Brisbane, QSL, Australia.

Lehman, P. (1968). *Tests and Measurements in Music.* Englewood Cliffs, N. J.: Prentice-Hall.

Lemon, N. (1973). *Attitudes and their measurement.* London: Batsford.

Lyman, H. B. (1971). *Test Scores and What They Mean.* Englewood Cliffs, New Jersey: Prentice-Hall.

Mager, R. (1975). *Preparing Instructional Objectives* (2nd ed.) Belmont, CA: Fearon-Pitman Publishers, Inc.

Mayhew, L. B. (1958). And in Attitudes. In Dressel, P. L. (Ed.) *Evaluation in the Basic College at Michigan State University,* New York: Harper.

McGuire, W. J. (1969). *The nature of attitude and attitude change.* In Linzey, G. & Aronson, E. (Eds.), The handbook of social psychology (Vol. 3). Reading, Mass.: Addison-Wesley.

Minato, S., & Kamada, T. (1996). Results of Research Studies on Causal Predominance between Achievement and Attitude in Junior High School Mathematics of Japan. *Journal for Research in Mathematics Education*, *27*(1). 96-99.

Ministry of Education, EARU. (1997) NEMP Music Assessment Results 1996. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L.

Ministry of Education, EARU. (2001) NEMP Music Assessment Results 2000. (*National Education Monitoring Report 4.).* Dunedin, NZ.: Crooks, T. and Flockton. L.

Ministry of Education, EARU. (2005) NEMP Music Assessment Results 2004. (*National Education Monitoring Report 32.).* Dunedin, NZ.: Crooks, T. and Flockton. L.

Ministry of Education. (2000). *The Arts in the New Zealand Curriculum.* Wellington, N.Z.: Author.

Mittler, G. (1972). Efforts to secure congruent and incongruent modifications of attitude toward works of art. *Studies in Art Education, 13*(2). 58-70.

Morris, J. W., & Stuckhardt, M. H. (1977). Art Attitude: Conceptualisation and Implication. *Studies in Art Education. 19*(1). 21-28.

Mueller, K. H.(1956). Studies in Music Appreciation, *Journal of Research in Music Education, 4*(1). 3-25.

Mursell, J.L. (1937). *The Psychology of Music.* New York, W.W. Norton and Co.

National Academy of Education. (1993). Setting performance standards for student achievement. *U.S. General Accounting Office*. Stanford, CA: Author.

New Zealand Ministry of Education. (1993). *New Zealand Curriculum Framework,* Wellington, N.Z.: Author.

New Zealand. *The New Zealand Curriculum Framework.* (1993). Ministry of Education. Wellington. Learning Media.

Popham, W. J. (1978) *Criterion-referenced Meausrement.* Englewood Cliffs, NJ: Prentice-Hall.

Resnick, B. (2004). Majority of districts/schools employ "high-stakes" testing. *Successful School Marketer*. Retrieved December 9, 2005, from http://www.schooldata.com/ssm-resnick-majority.htm

Seashore, C. E. (1938). *Psychology of Music.* New York, McGraw-Hill.

Schneider, E. H. & Cady, H. L. (1965). Evaluation and Synthesis of Research Studies Relating to Music Education. *Cooperative Research Project. E-016*, Office of Education, U.S. Department of Health, Education and Welfare. Columbus, Ohio: The Ohio State University Research Foundation.

Shuter-Dyson, R. (1968). *The Psychology of Musical Ability.* London: Methuen & Co Ltd.

Shuter-Dyson, R., & Gabriel, C. (1981). *The Psychology of Musical Ability* (2nd edi.). London: Methuen.

Stuckhardt, M. H. (1973). A critical review of attitude scales used in the visual arts. *Review of Research in Visual and Environmental Education. 1*(2). 7-23.

Thurstone, L. L. (1928). Attitudes Can Be Measured. *The American Journal of Sociology*, *33*(4). 529-554.

Tinkham, T. (1989). Rote Learning, Attitudes, and Abilities: A Comparison of Japanese and American Student. *TESOL Quarterly*, *23*(4). 695-698.

Tyler, L. E. (1946). An Exploratory Study of Discrimination of composer Style, *The Journal of General Psychology,* 34.

Tyler, R. W. (1966). The objectives and plans for a National Assessment of Educational Progress. *Journal of Educational Measurement, 3*(1). 1-4.

Wing, H. (1948) *Tests of Musical Ability and Appreciation.* London: Cambridge University Press.

Wing, H. (1948). *Standardised Tests of Musical Intelligence.* National Foundation for Educational Research. Upton Park, Slough.